BTP Report

Abstract:
Analysing neuron structure is critical for understanding how they function within neural circuits. Neuron tracing or Neuron reconstruction is a technique used in computational neuroscience to determine the path of neural axons and dendrites from advanced microscopic images. Due to the high complexity of neuron morphology and often seen heavy noise in such images, as well as the typically encountered massive amount of image data, it has been widely viewed as one of the most challenging computational tasks for computational neuroscience. The goal of this project is to develop a completely automatic method for tracing the neuron morphology in 3D.

Neuron reconstruction is critical to understanding the neuron anatomy. By understanding the single neuron anatomy, it becomes easier to understand the how a particular neuron is connected to other neurons. This also helps us to better understand certain properties like cell-firing etc. Also, knowing the right morphology of the neuron helps us to analyse what changes are seen in its structure with age and certain conditions like stress or disease etc.

Acknowledgement:
We would like to acknowledge our B. Tech. project advisor Dr. Deepti Bathula for guiding us through her expertise over medical image processing. We would also like to thank Ms. Apoorva Sikka for at times helping us out with our doubts and queries. As team members, we thank each other too for each one's immense contribution to the project.

Honour code:
We hereby certify that, we have cited, obtained permission statement from the owner(s) of any third party copyrighted matter to be included in our project report. We also take full responsibility of the submitted project code and contents of this report.

Shubham Sharma (2013CSB1114)

Nitin Kumar (2013CSB)
Saurabh Khoria (2013CSB1029)

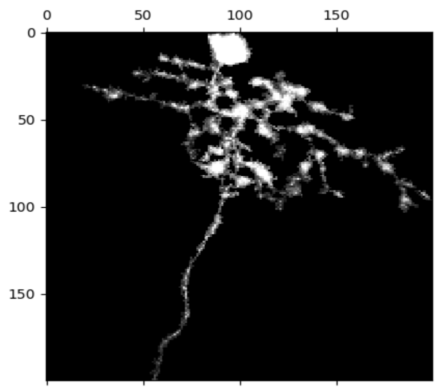
Certificate

It is certified that the work contained in this report titled “Single neuron reconstruction” has been done by Shubham Sharma(2013CSB1114), Nitin Kumar(2013CSB1029) and Saurabh Khoria(2013CSB1029) and has been carried out under my supervision.

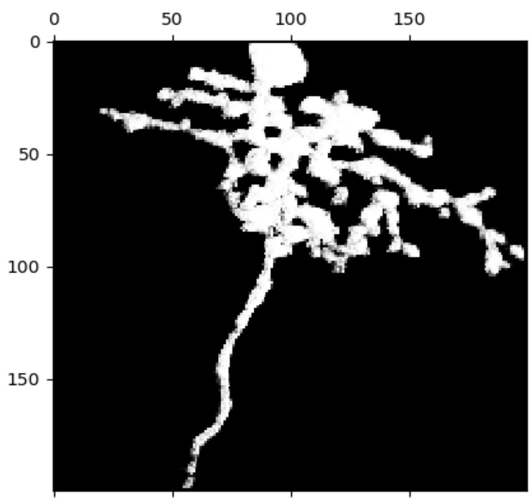
Dr. Deepti Bathula
Project Supervisor
Department of Computer Science & Engineering
Indian Institute of Technology Ropar
Rupnagar-140001

Contents:

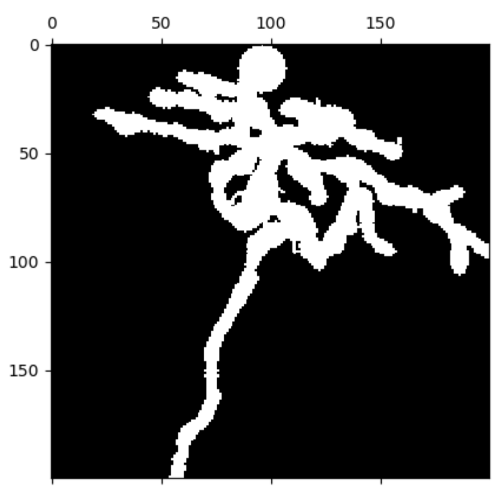
List of figures:


0
50
100
150
0
50
100
150

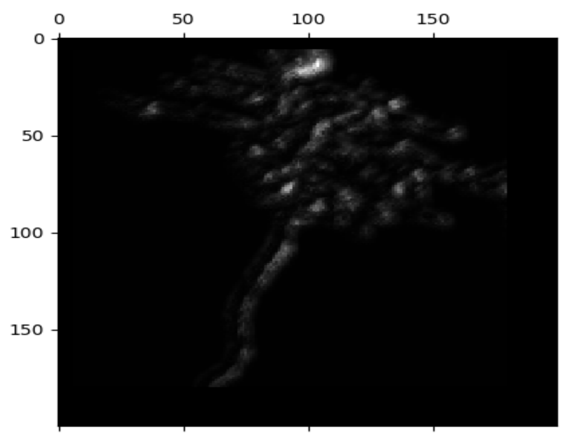
1.2


0
50
100
150
0
50
100
150

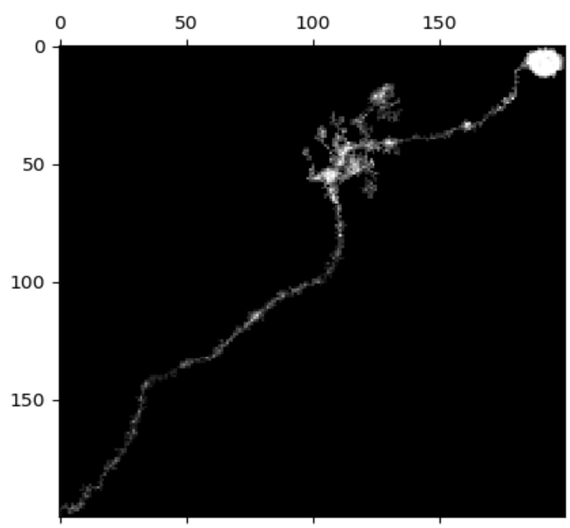
1.3


0
50
100
150
0
50
100
150

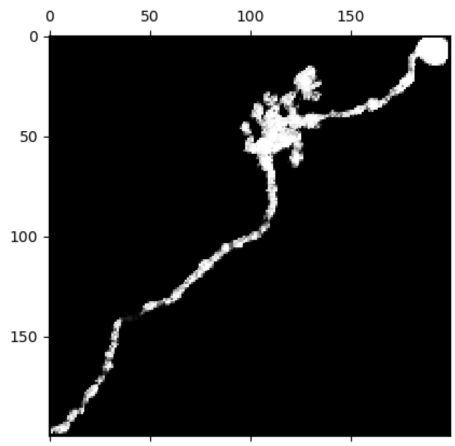
1.4



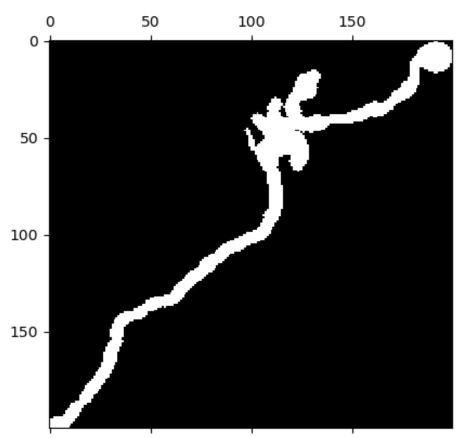
2.1



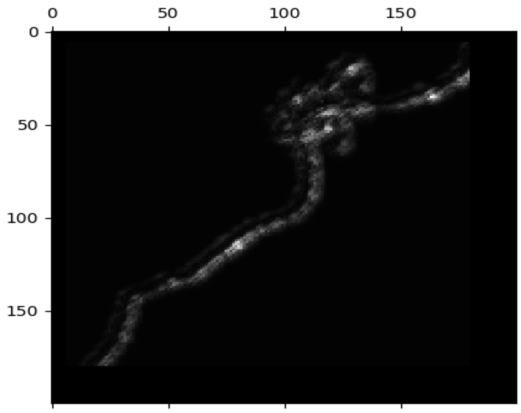
2.2



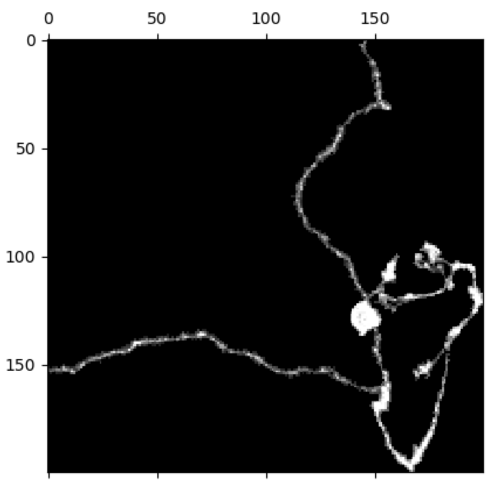
2.3



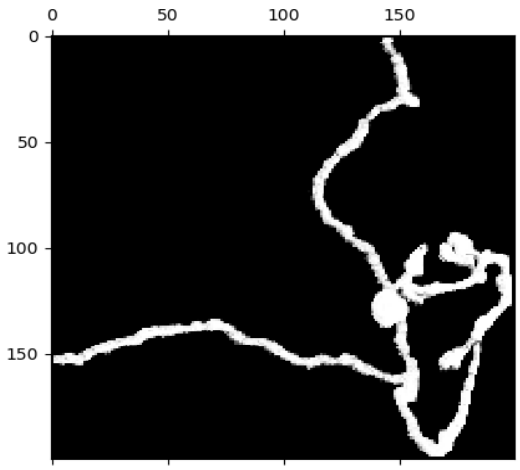
2.4



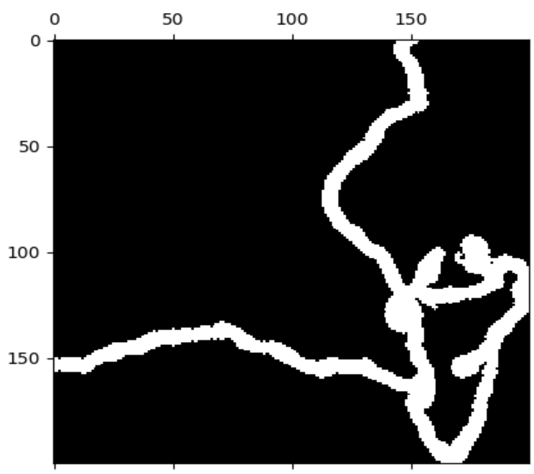
3.1



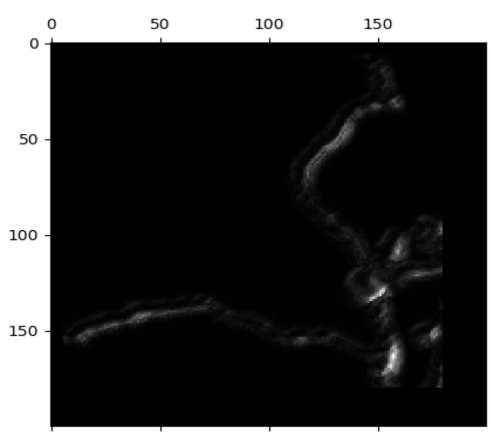
3.2


0
50
100
150
0
50
100
150

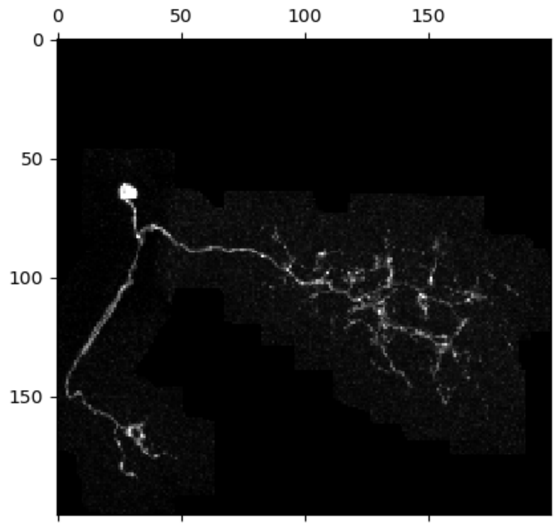
3.3


0
50
100
150
0
50
100
150

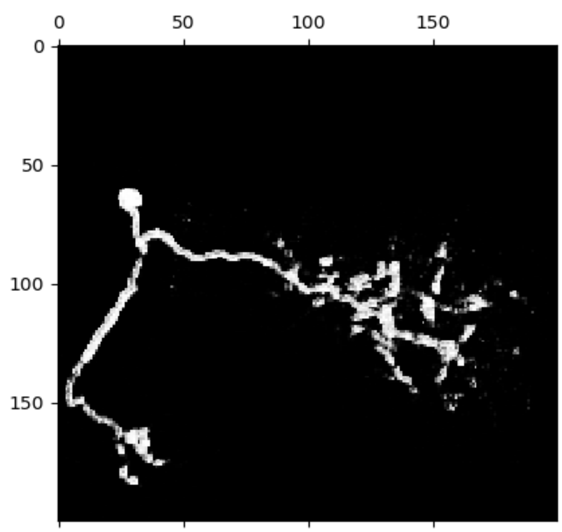
3.4


0
50
100
150
0
50
100
150

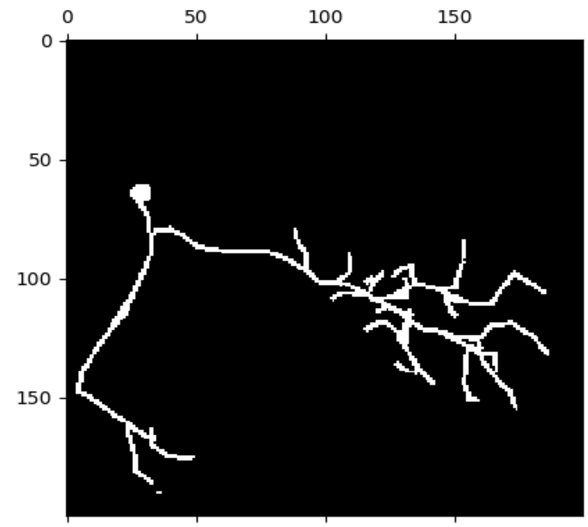
4.1



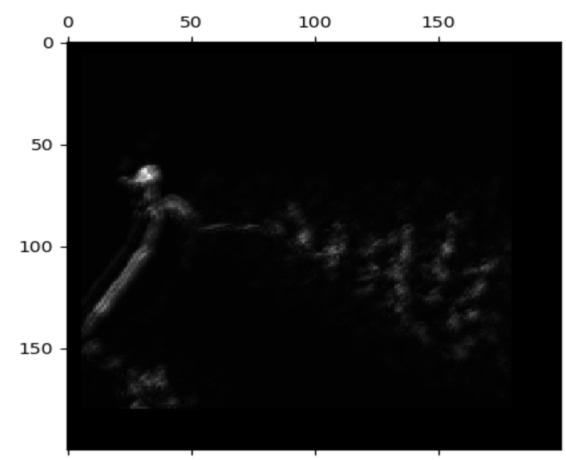
4.2



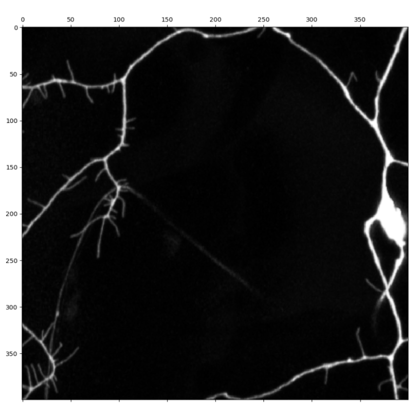
4.3



4.4

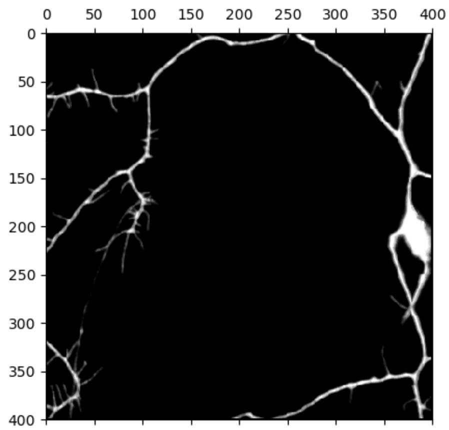


5.1



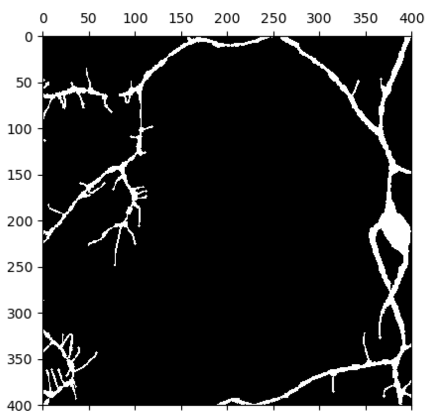
Original

5.2



Reconstructed

5.3

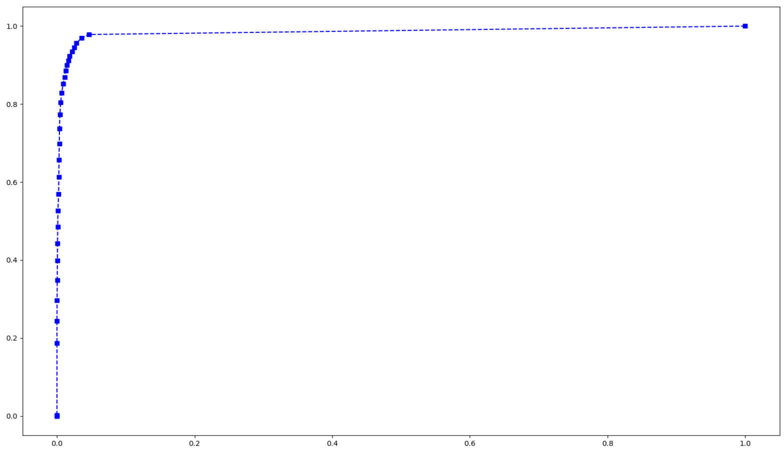


Ground truth

5.4

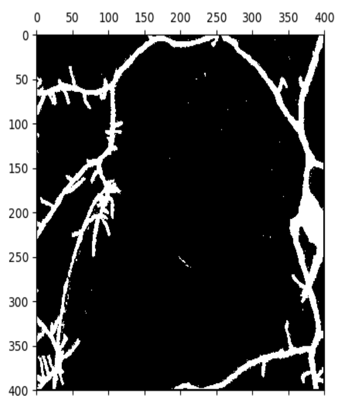
All path pruning algorithm

6.1

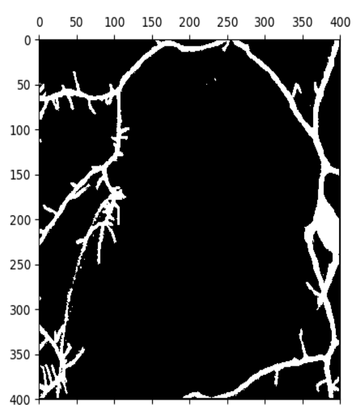


ROC fruitfly 7 equal patches

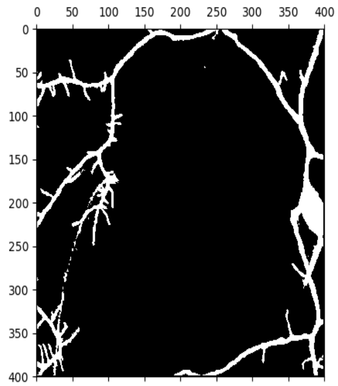
7.1



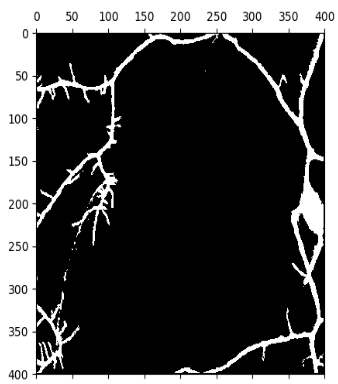
7.2



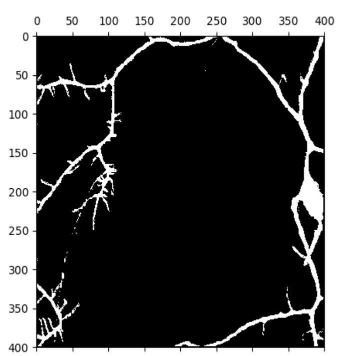
7.3


0 50 100 150 200 250 300 350 400
0 50 100 150 200 250 300 350 400

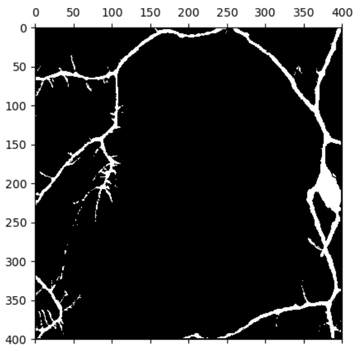
7.4


0 50 100 150 200 250 300 350 400
0 50 100 150 200 250 300 350 400

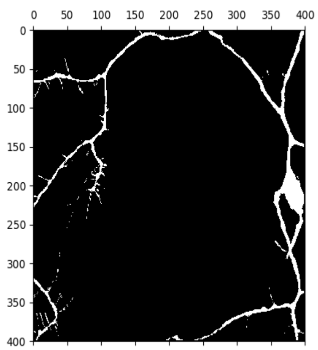
7.5


0 50 100 150 200 250 300 350 400
0 50 100 150 200 250 300 350 400

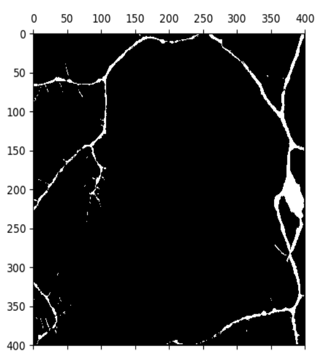
7.6


0
50
100
150
200
250
300
350
400

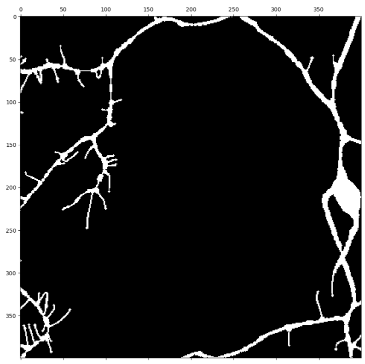
7.7


0
50
100
150
200
250
300
350
400

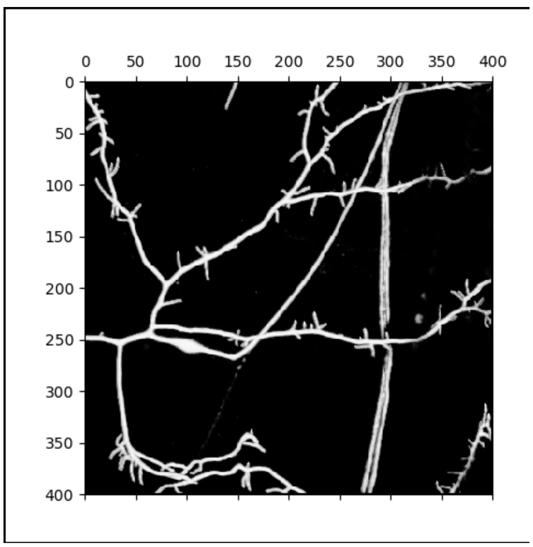
7.8


0
50
100
150
200
250
300
350
400

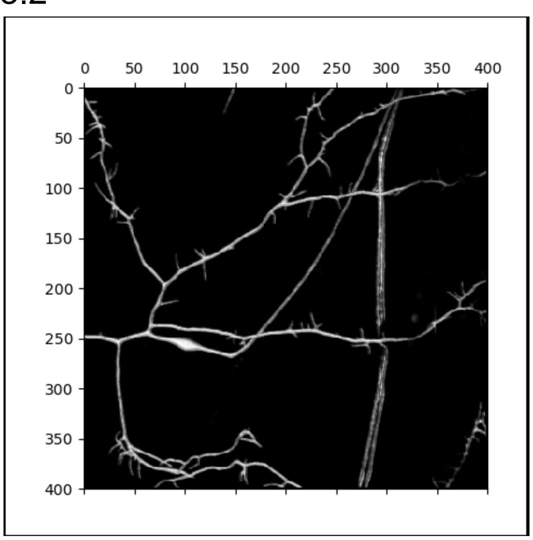
7.9



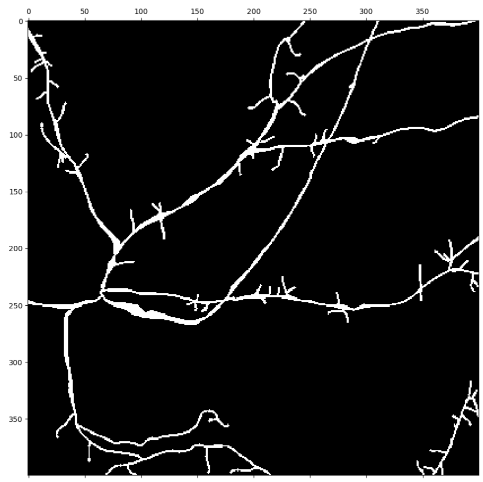
8.1


0 50 100 150 200 250 300 350 400
0 50 100 150 200 250 300 350 400

8.2

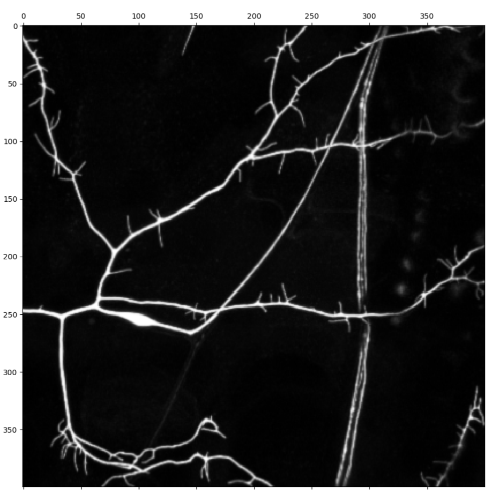

0 50 100 150 200 250 300 350 400
0 50 100 150 200 250 300 350 400

8.3



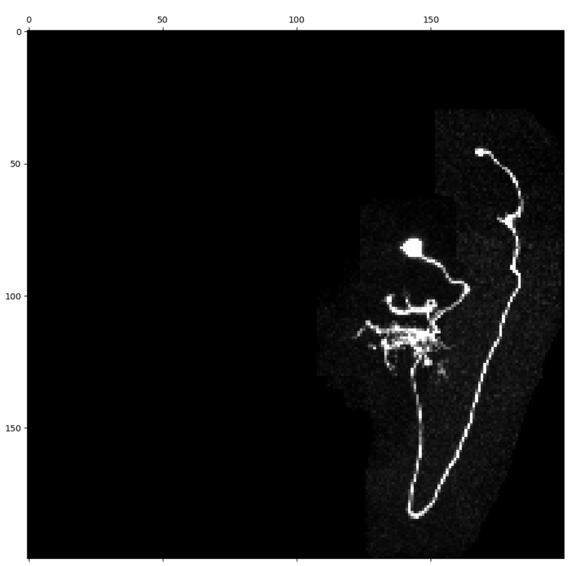
0
50
100
150
200
250
300
350

8.4



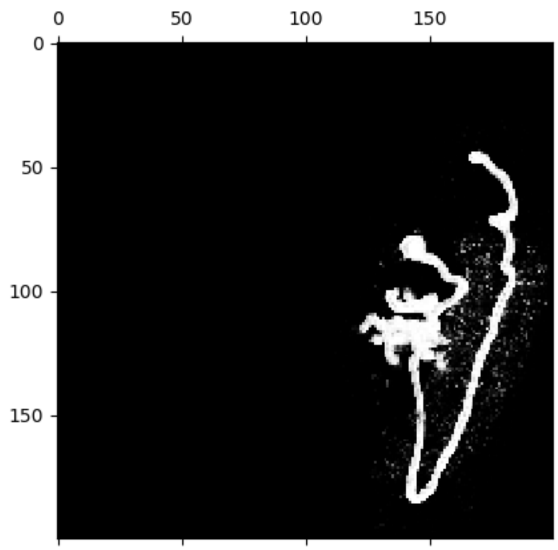
0
50
100
150
200
250
300
350

9.1

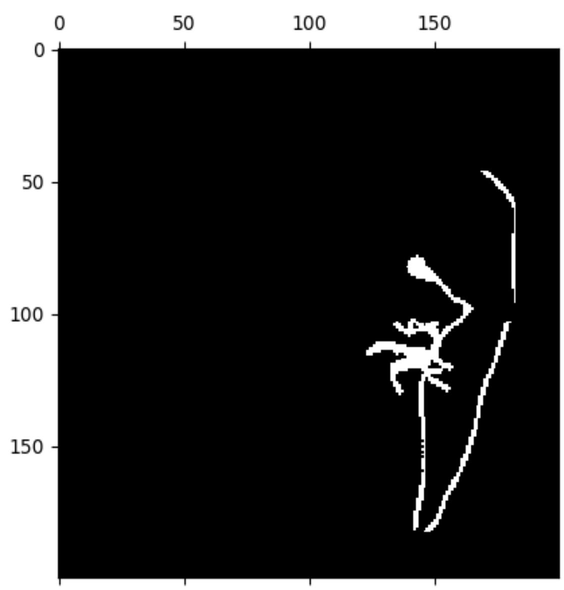


0
50
100
150

9.2

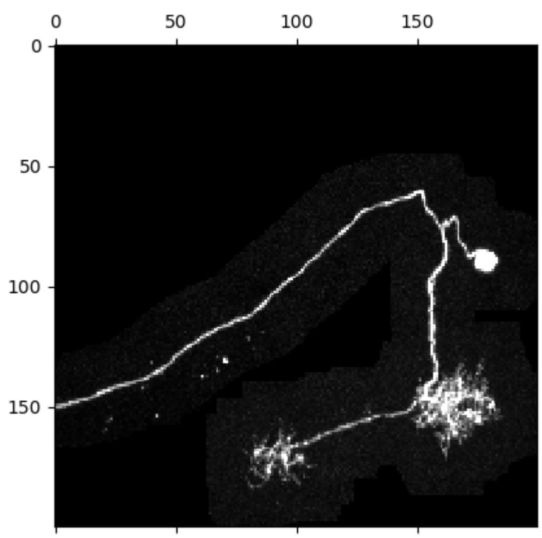


9.3

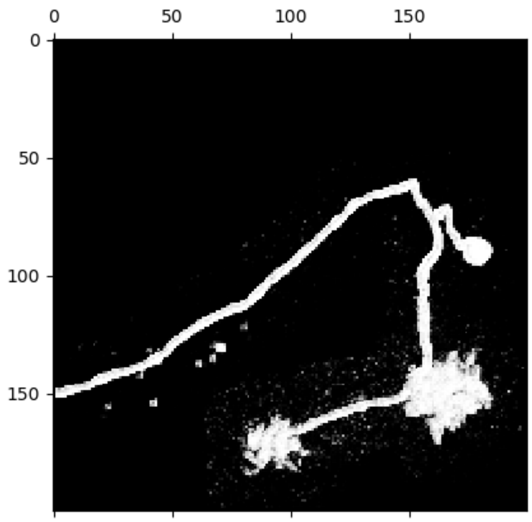


9.4

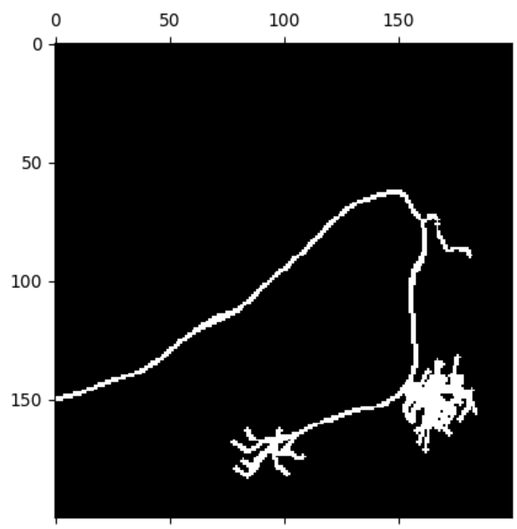
10.1



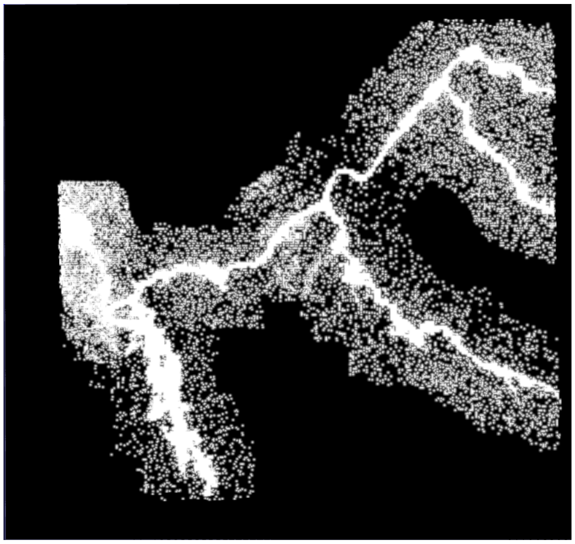
10.2



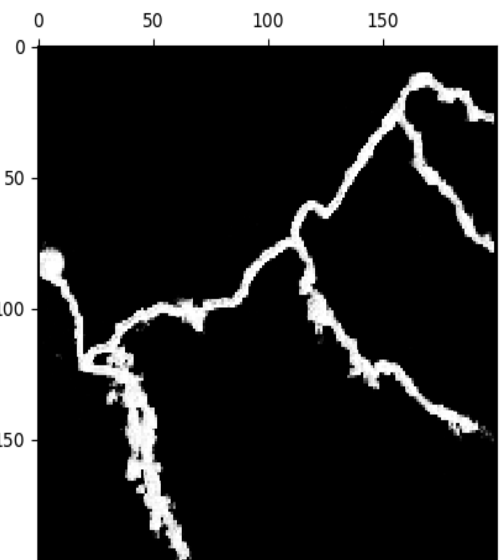
10.3



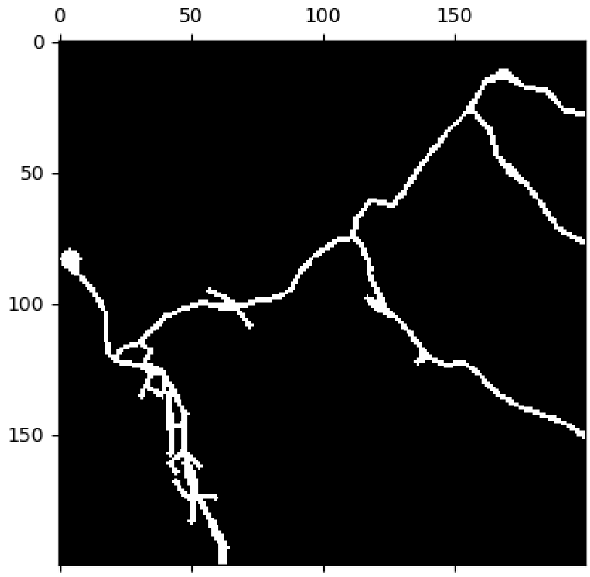
11.1



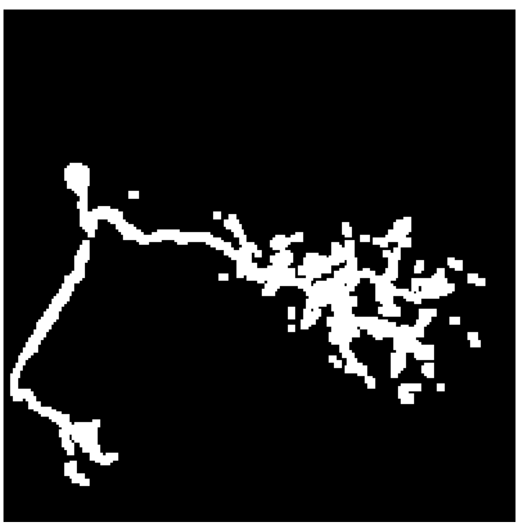
11.2



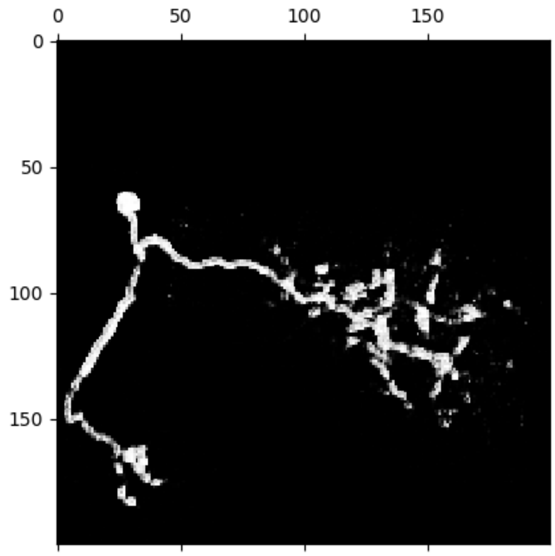
11.3



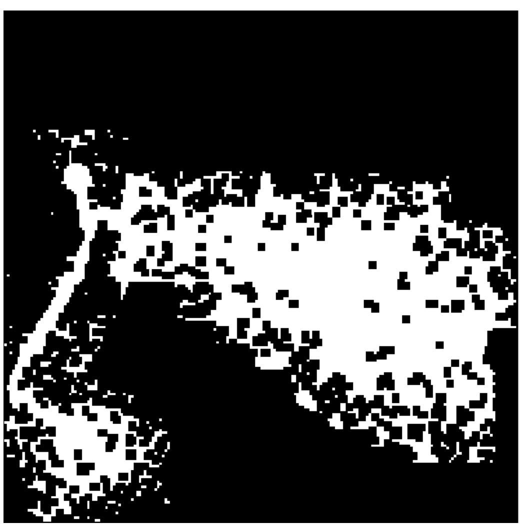
12.1



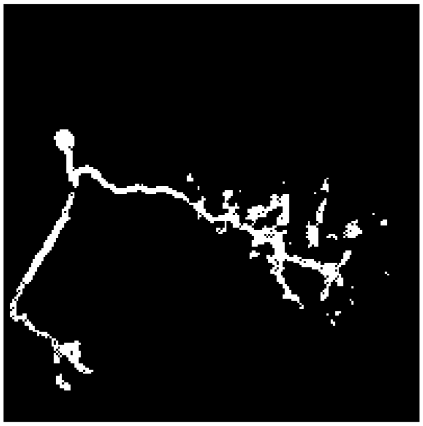
12.2



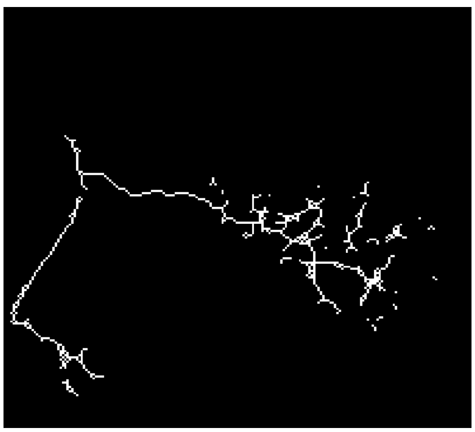
12.3



13.1

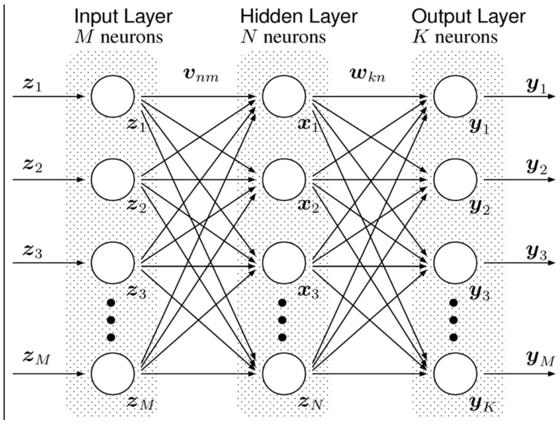


13.2

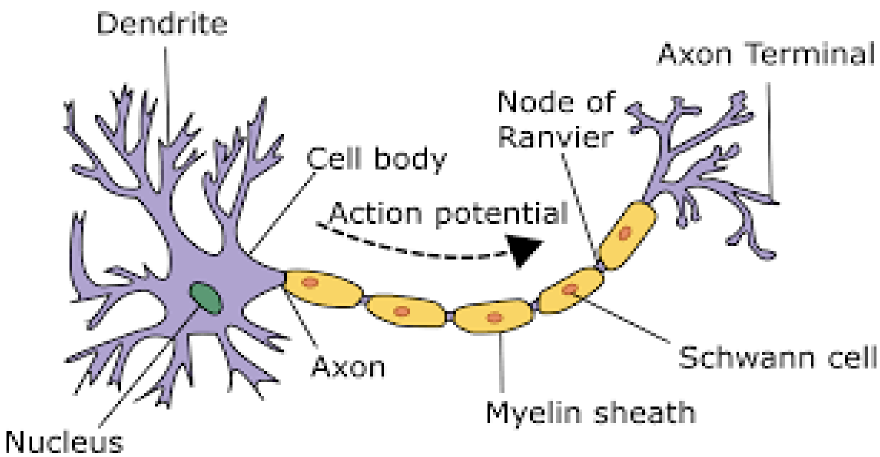


13.3

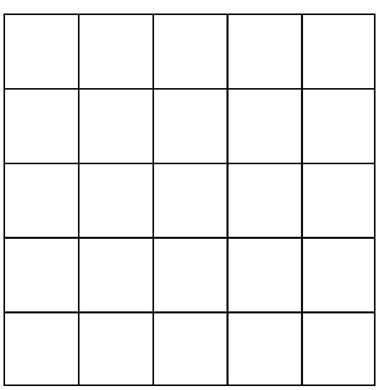
14.1


Input Layer
M neurons
Hidden Layer
N neurons
Output Layer
K neurons
v_{nm}
w_{kn}
z_1
z_2
z_3
z_M
x_1
x_2
x_3
z_N
y_1
y_2
y_3
y_M
y_K

15.1

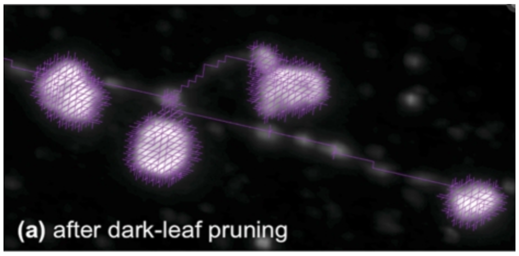

Dendrite
Axon Terminal
Node of
Ranvier
Cell body
Action potential
Axon
Schwann cell
Myelin sheath
Nucleus

15.2

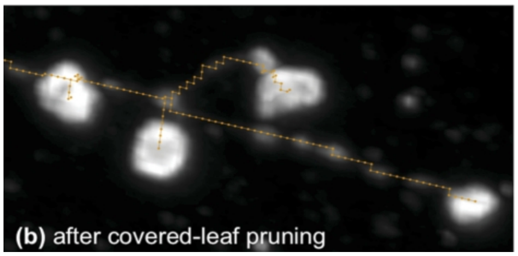


15.3

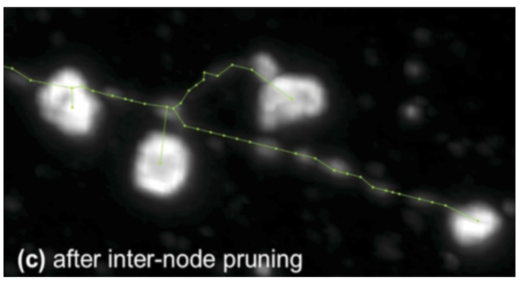
16.1



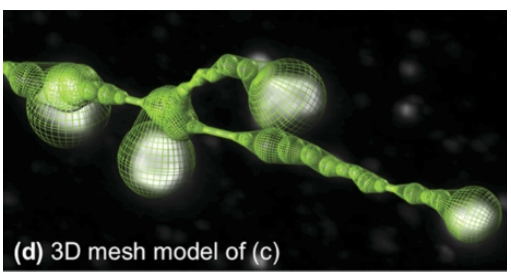
16.2



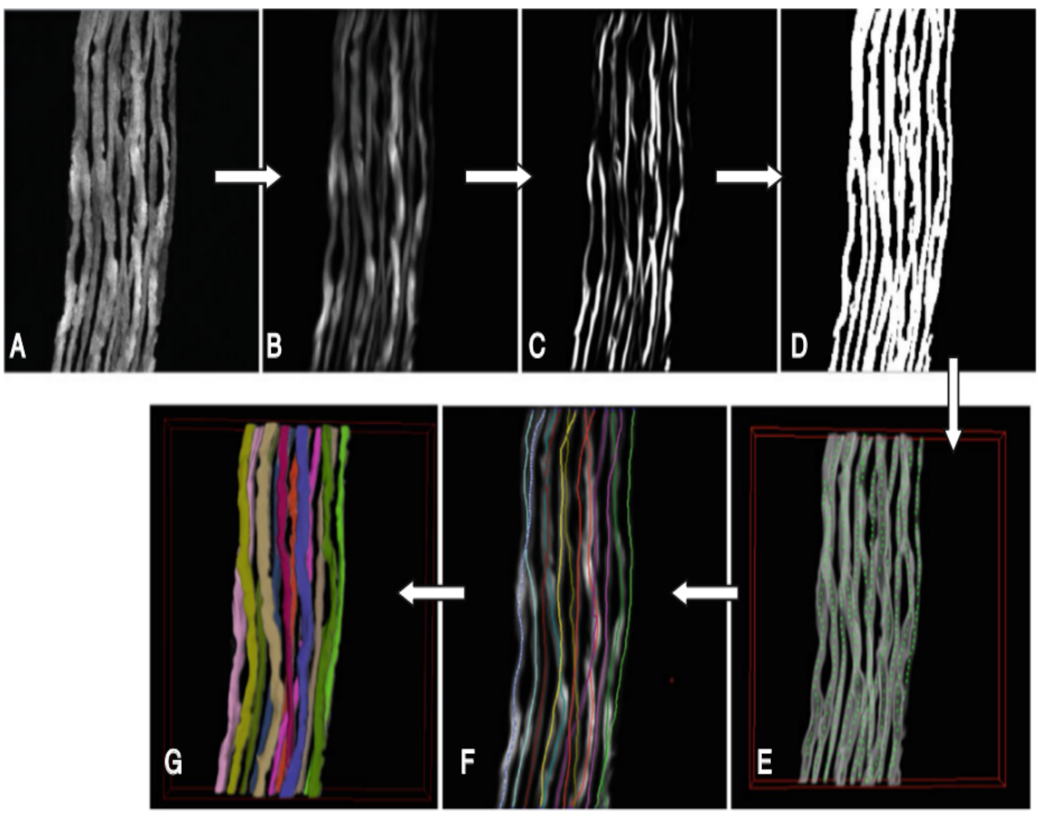
16.3



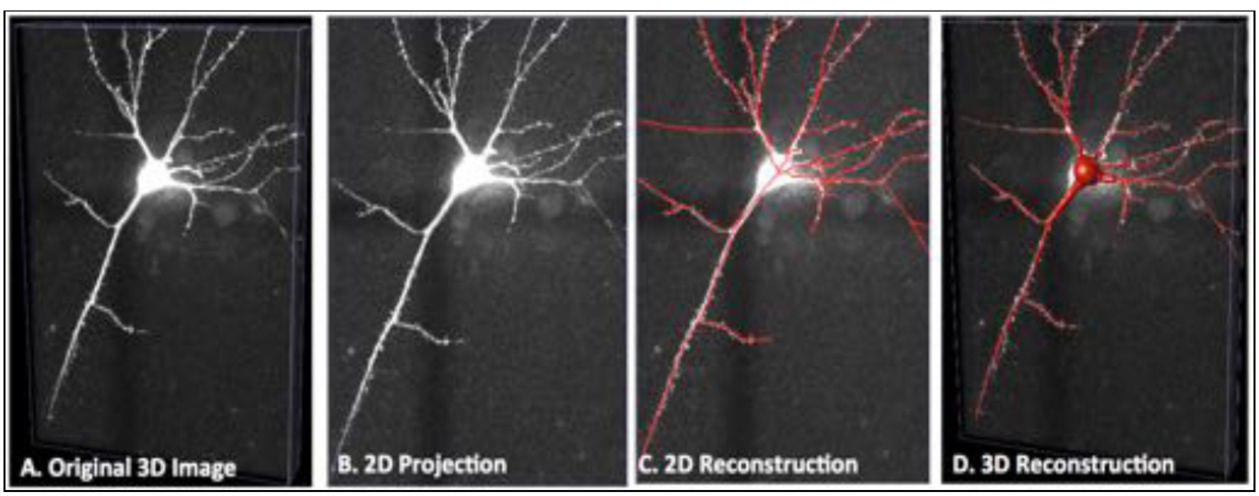
16.4



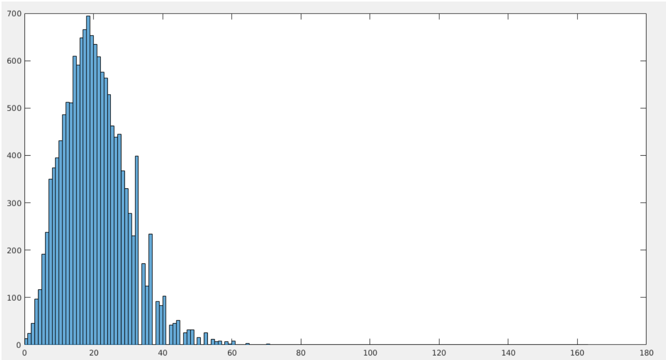
17.1


A
B
C
D
G
F
E

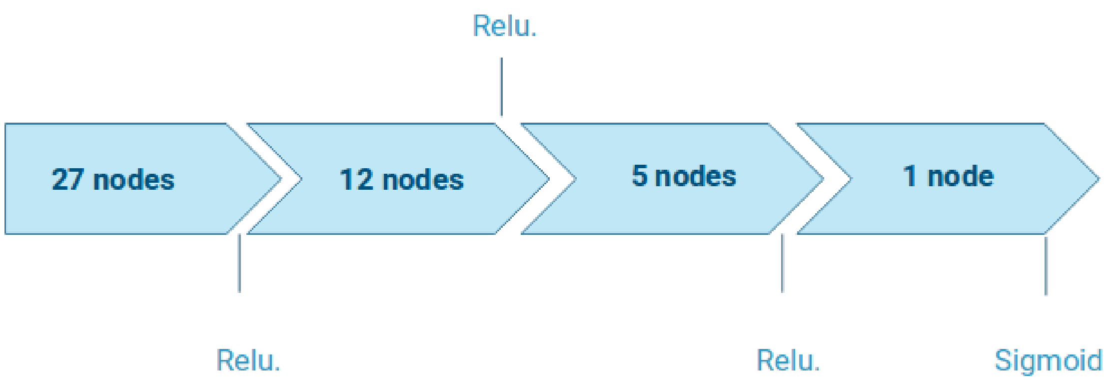
18.1


A. Original 3D Image
B. 2D Projection
C. 2D Reconstruction
D. 3D Reconstruction

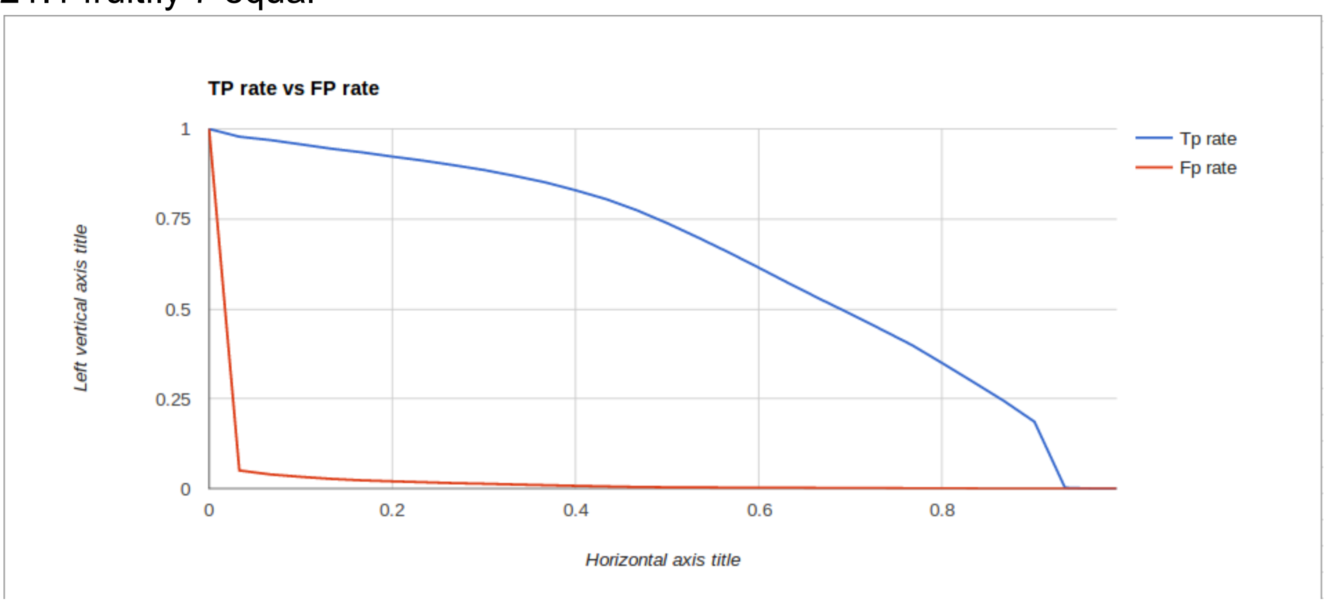
19.1



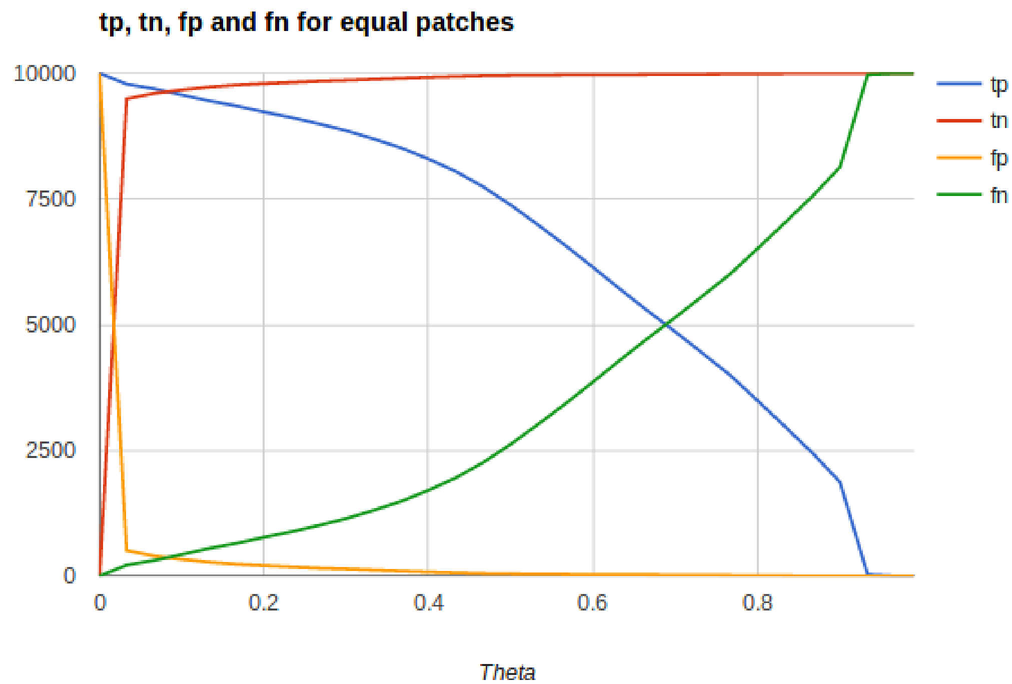
20.1



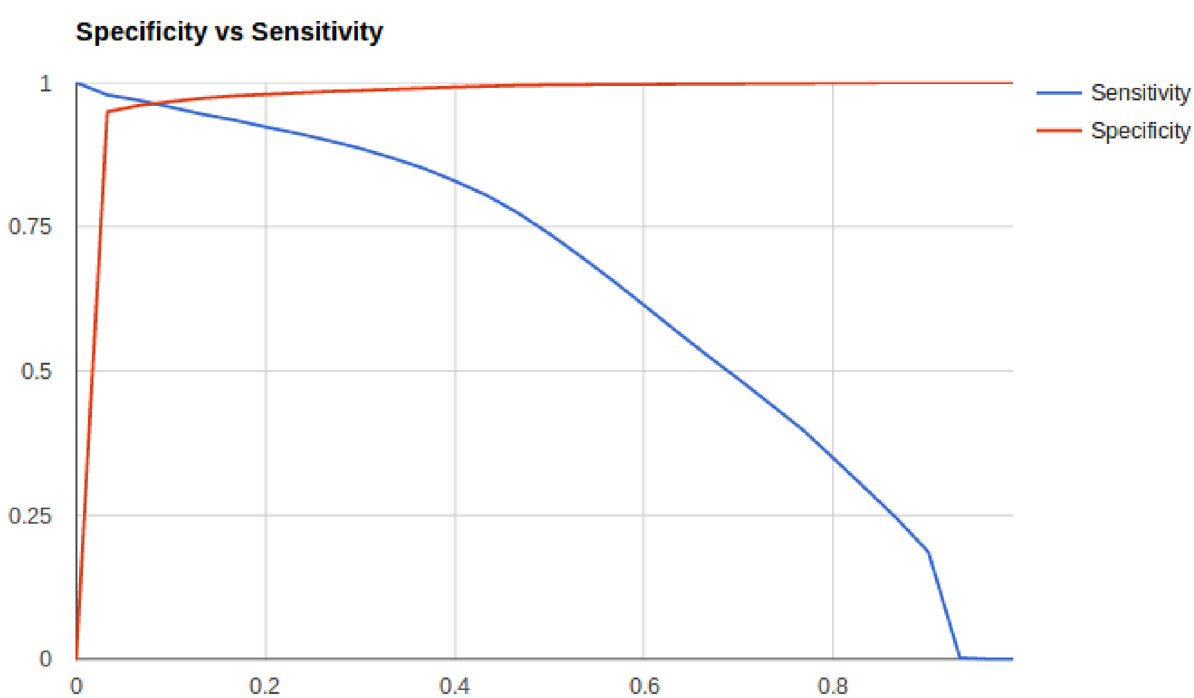
21.1 fruitfly 7 equal



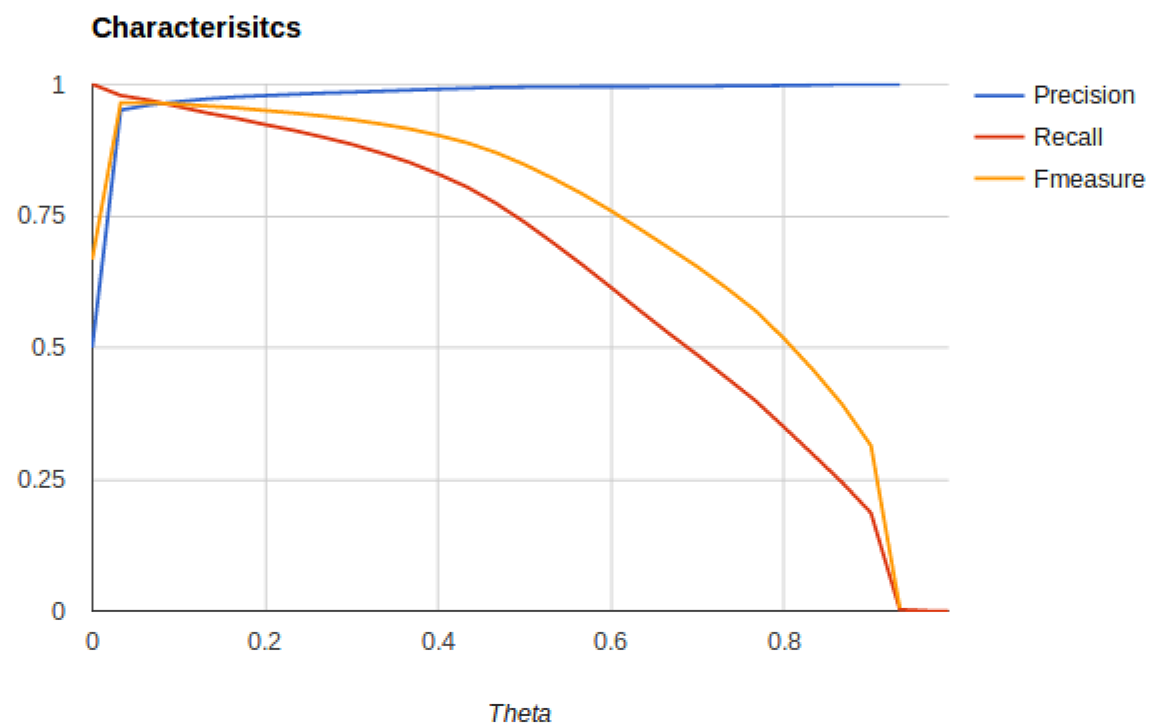
21.2 fruitfly 7 equal



21.3 fruitfly 7 equal



21.4 fruitfly 7 equal



21.5 fruitfly equal

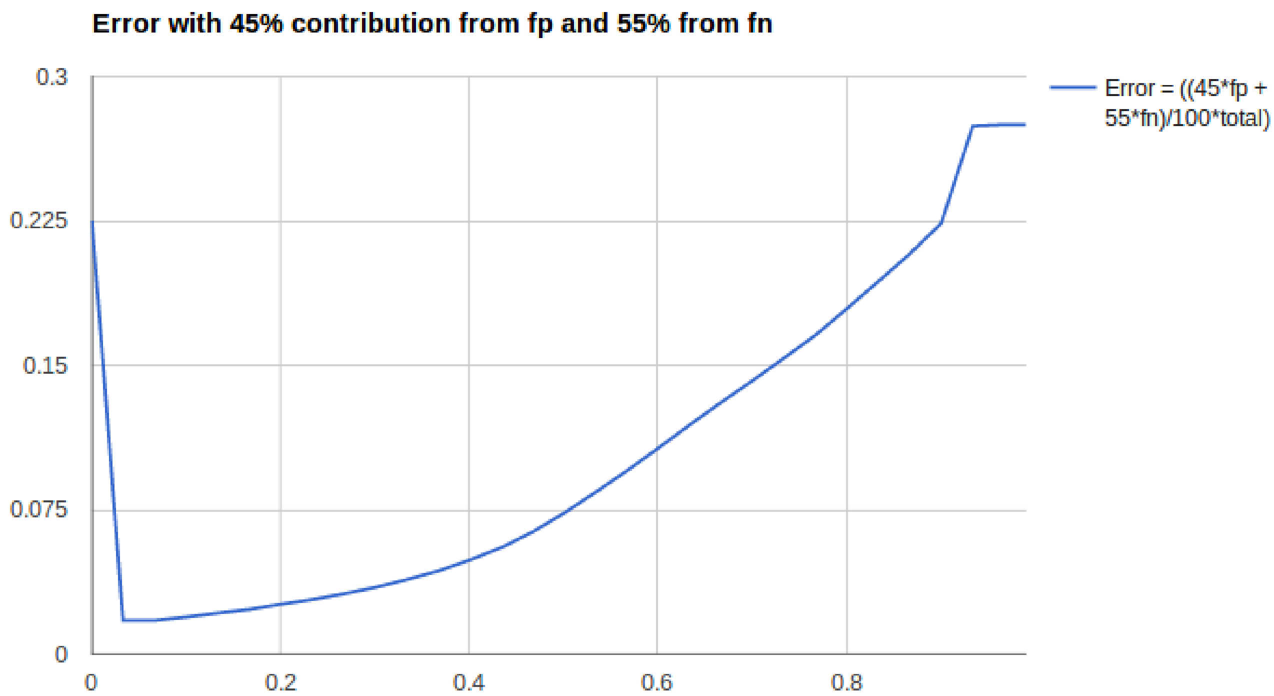
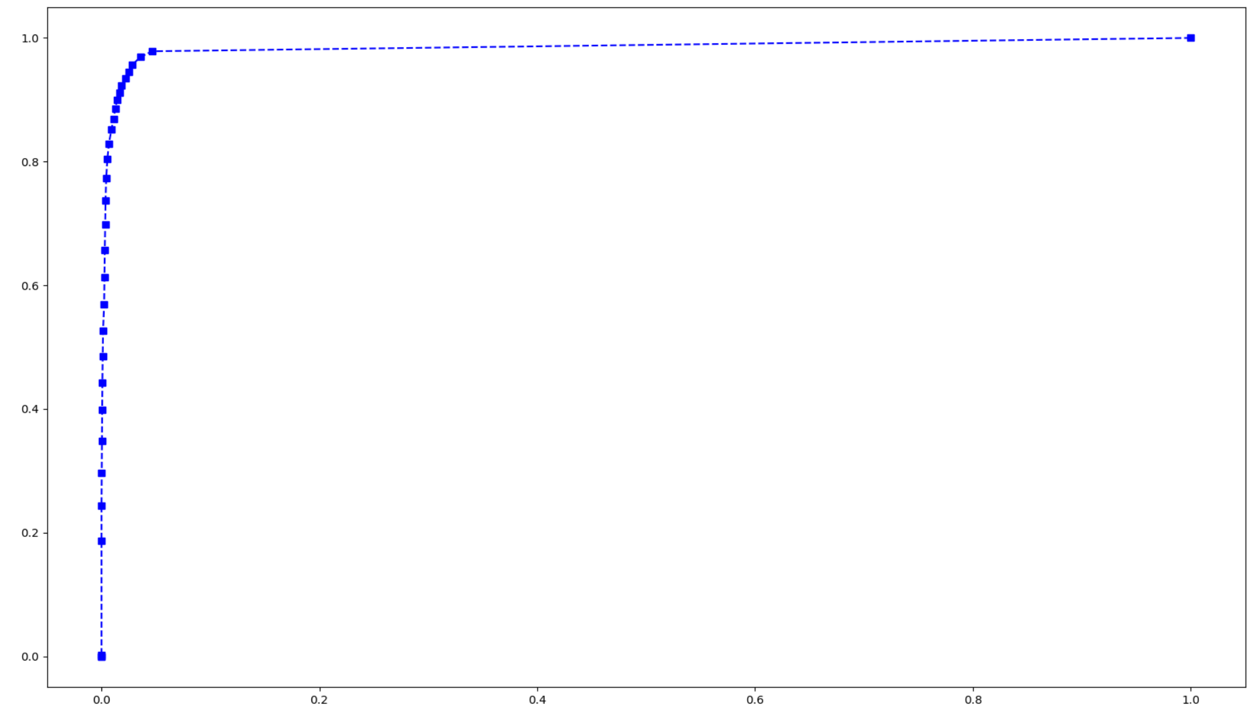

Error with 45% contribution from fp and 55% from fn
0.3
0.225
0.15
0.075
0
0
0.2
0.4
0.6
0.8
Error = ((45*fp + 55*fn)/100*total)

Fig 21.6



1. Introduction to neuron tracing and contemporary methods

Digital neuron tracing is critical to understanding the structure of the neuron. But in most of the images obtained from the 3D microscopic data, the signal to noise ratio is low and the neuron structure is also fragmented. Manually annotated neuron structure is possible but is time consuming and cumbersome. The aim is to come up with a fully automatic algorithm which can perform the 3D reconstruction of neuron morphology.

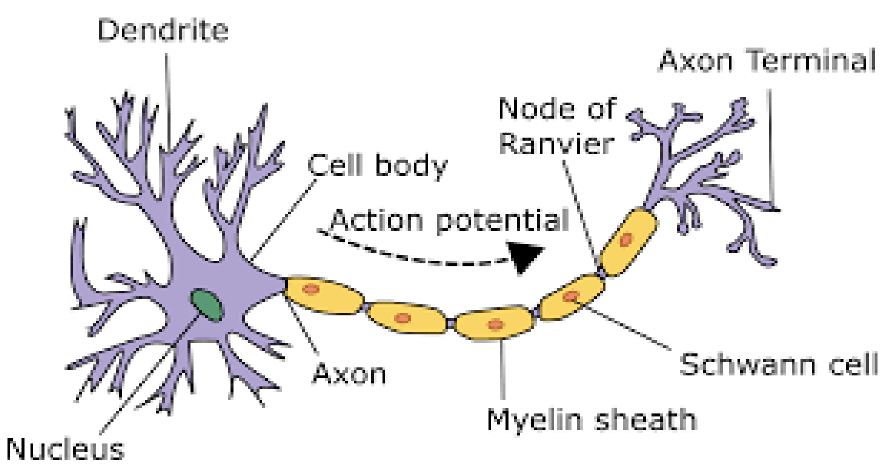


Fig 15.1

The main state-of-the-art algorithms used in 3D neuron reconstruction are:

1. All path pruning
2. Treemap algorithm
3. Open snake curve algorithm

1.1 All path pruning

This algorithm first produces an initial over-construction of the neuron skeleton. Then, the neuron structure is pruned to remove the redundant nodes. It is assumed that we can detect the soma of the neuron automatically, owing to the fact that it is one of the big and bright spots in the image. Thresholding is done on the voxels using the mean intensity of the image which we call t_a. One can optionally run 3*3*3 median filter to remove noise.

The algorithm takes a neuron image 3D and a seed location Ps = (xs, ys, zs). The seed is often the soma of the neuron. Neuron signal is assumed to be represented by bright but not dark image voxels. The weight of the edge between two voxels at two vertices v0 = (x0, y0, z0) a+nd v1 = (x1, y1, z1) and is given by:

$$e(v0, v1) = ||v_0 - v_1|| (g_t(v_0) - g_t(v_1))/2$$

$$g_t(p) = \exp(\lambda_l(1-I(p)/I_{max}))$$

I_{max} refers to the max intensity in the image

All path pruning uses dijkstra’s algorithm to find the shortest path from location P_s to every other vertex in G. Since this also gives us the shortest path of every node from P_s, only those vertices are included in the reconstruction which appear in the shortest path to P_s of some neuron.

This graph is highly sparse because the edges are between voxels which are nearby. Time complexity to solve the one source shortest path problem is

O(NlogN).

Pruning:

The next job is to prune the redundant nodes from the reconstruction so that only the necessary voxels remain in the final image. For doing this, the following steps are performed in order:

1. Dark leaf pruning
2. Covered leaf pruning
3. Inter-node pruning

Dark leaf pruning: In dark leaf pruning, we choose the average intensity of the voxels in the image t_a or the lowest visible intensity t_v in the image and use this to threshold the image.

Covered leaf pruning:

Define a radius adjustable sphere centered at reconstruction node, and enlarge the radius gradually until 0.1% of the image voxels within the sphere are darker than the global threshold t_a which is the average voxel intensity of the entire image.

For two reconstruction nodes a and b, a is significantly covered by b if they satisfy the following conditions:

$\Omega(a \cap b)/\Omega(a) > 0.9$

When no node covers the leaf node, the leaf node needs to be kept.

If a leaf node is covered by another node, the node can be safely pruned.

For each node a, create an empty covering list C_a, which would store the identities of other nodes that cover node a.

While going through all the nodes, check if that node covers a or not. If yes, put that in the covering list of a.

Inter-node pruning:

The reconstruction can further be improved by the following steps:

Start from a leaf node a, for its immediate parent b, check if b is significantly covered by a.

$$\Omega(a, b)/\Omega(b) > 0.1$$

If the condition is met, remove b. Make parent of b the current parent of a.

Do that until you encounter a branch node or a root node.

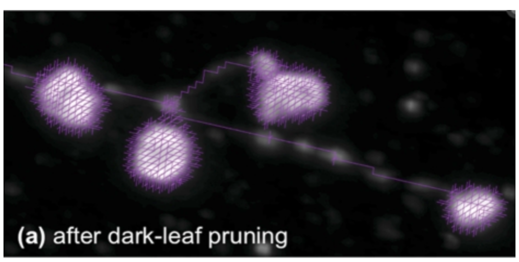


Fig 16.1

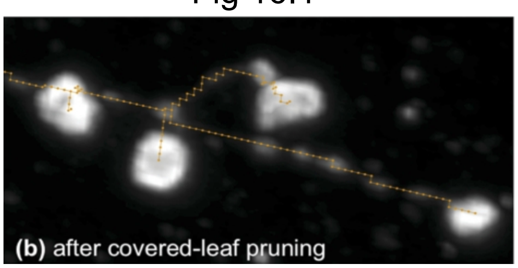


Fig 16.2

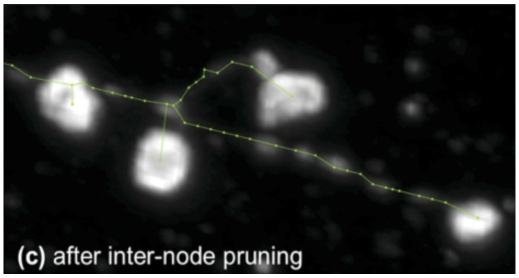


Fig 16.3

The all path pruning removes most of the image voxels from the 3D image. In the images cited above ie Fig 16, out of 94741 image voxels only 418 remain in the end. The all path pruning algorithm is fast and quite robust to choice of location of seeds and image noise.

1.2 TReMAP

TReMAP ie. Tracing, Reverse Mapping and Assembling of 2D Projections , utilizes 3D Virtual Finger (a reverse-mapping technique) to detect 3D neuron structures based on tracing results on 2D projection planes using APP algorithm.

Instead of tracing a 3D image directly in the 3D space as seen in majority of the tracing methods, we first trace the 2D projection trees in 2D planes, followed by reverse-mapping the resulting 2D tracing results back into the 3D space as 3D curves.
then we use a minimal spanning tree (MST) method to assemble all the 3D curves to generate the final 3D reconstruction
We simplify a 3D reconstruction problem into 2D, the computational costs are reduced dramatically.

Algorithm

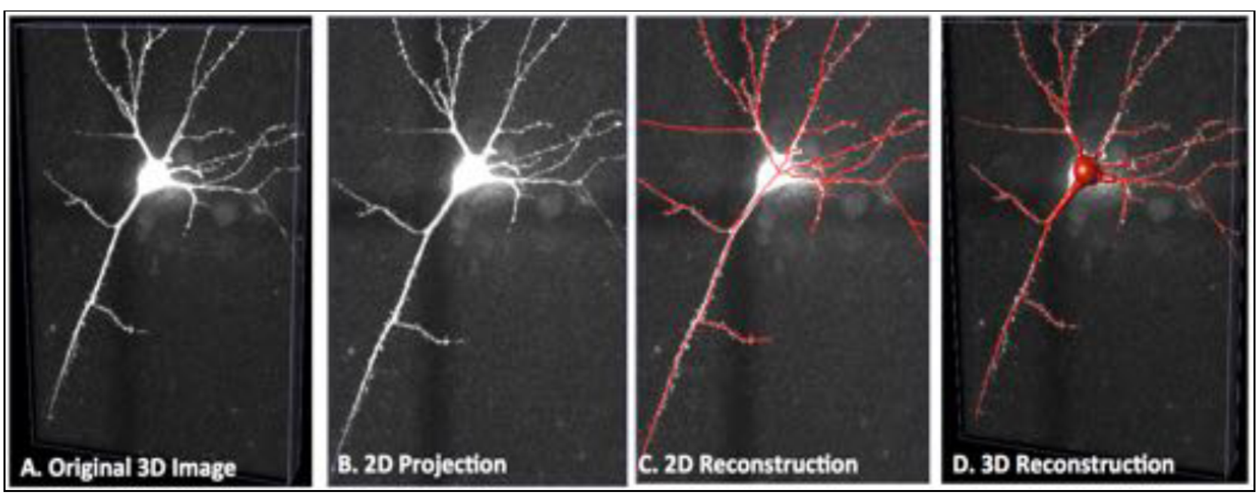
- Generate the 2D maximum intensity projections (MIPs) from an input 3D image
- For each 2D MIP image, the foreground and background are separated based on intensity thresholding.
- In foreground, all connected components are identified and labeled with different indexes.

Tracing the 2d Projection tree

We trace the signals in 2D projections (Figure 3) by using allpath-pruning tracing algorithms, which often produce a single neuron tree structure that has less false positives than most other tracing algorithms.
In the cases of all-path-pruning algorithms, a neuron is reconstructed via first searching all possible signals from a single neuron image to generate an over-reconstructed tree. Then, several different pruning strategies are used to simplify the neuron tree.

Reverse Mapping and Assembling

- Extract the 2D neuron tree by tracing a projection image.
- It is broken into a collection of 2D segments.
- a 2D bounding box is identified and used to extract the corresponding 3D image content for reverse mapping.
- divide-and-conquer strategy is crucial to cut down the memory consumption of reverse mapping and it enables parallel processing that can further cut down computation time.



How to Map

Curve Drawing Algorithm 2 is used from Virtual Finger Method
From each node of the 2D curve, a shooting ray is generated along orthogonal direction of the 2D projection plane. The starting 3D node is identified by the maximal intensity location along the first shooting ray (at the first 2D node). Then CDA2 searches for 3D curve path from the last 3D node to the current shooting ray using an

adapted fast-marching algorithm. This process iterates over all the 2D nodes and forms the corresponding 3D curve.

A 3D reconstructed neuron is often represented using a tree graph, which is made up of a series of reconstruction nodes. Each node has its own 3D X, Y, Z locations, radius, as well as the information which other node is this node's topological "parent." A neuron reconstruction can also be viewed as a connected graph of many "segments," each being a 3D curve consisting of a number of reconstruction nodes. In the case of TReMAP, each segment corresponds to a 3D curve produced using the Virtual Finger methods. In the end, MST is used to connect all 3D curves to produce the final neuron reconstruction.

Advantage :- Efficient computation (much less memory consumption and parallel computation) for large-scale images.

Chapter 2

Pre-processing and intuition for machine learning

The required data set of the microscopic images of neuron was obtained through the Allen brain institute which is one of the pioneering institutions in 3D neuron reconstruction. The main problem was with the format of the data which was in the form suitable to be read and interpreted in Vaa3d, a software where one can apply these existing neuron reconstruction algorithms and see and compare the outputs. The format of the data was not in the simple 3D matrix form which could be operated by our algorithms.

2.2 Pre-processing

It was observed that the neuron data of many species had a particular kind of noise called the Rayleigh noise.

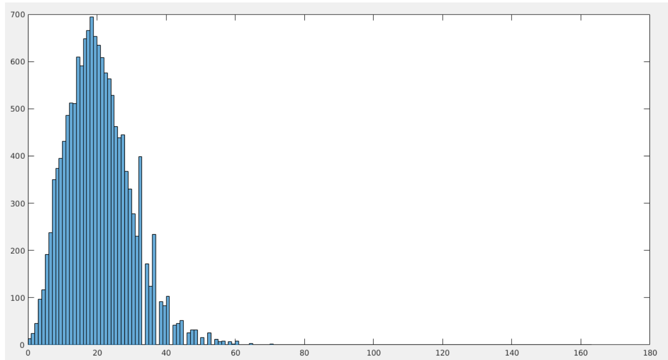


Figure 2.1

Anisotropic diffusion: Anisotropic diffusion is a technique for reducing noise in an image without affecting significant parts of the image like edges, lines or other details. This technique iteratively removes noise from the 3D images.

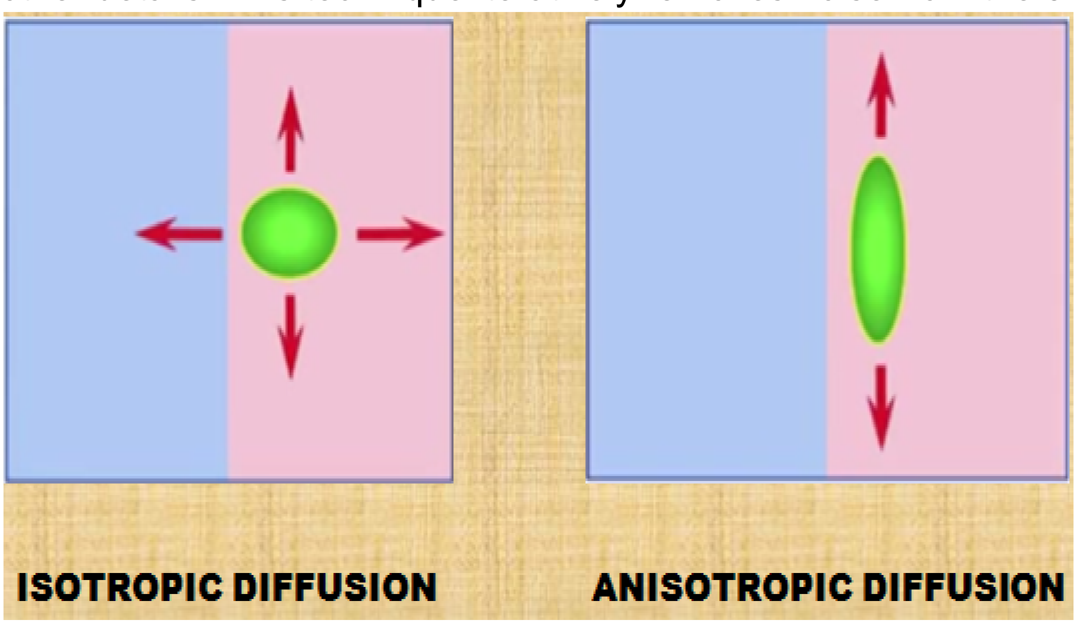
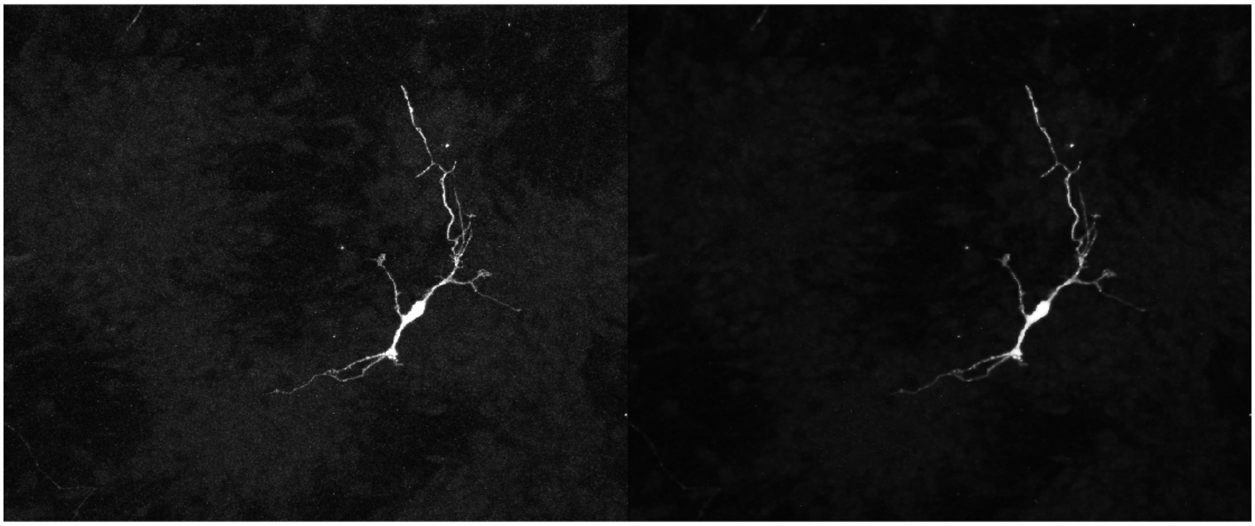


Figure 20.1



Original image Denoised image

Figure 2.3

2.3 Motivation for using machine learning

It was inferred that apart from using a predefined and rigid algorithm for tracing the structure of all the neurons in 3D, one can also look for a mechanism where our model or algorithm learns the structure of neuron by seen enough sample patches. This might also help us in cases where the structure of neuron varies across species and no one size fits all policy works. The learner or model can be trained with patches of neuron of a particular kind/specie or even with patches from across species and then the outcomes can be evaluated and compared.

The following kind of machine learning techniques were experimented with:

1. Simple neural network or multi-layer perceptron
2. Convolutional neural network

Chapter 3 Multiple Layer Perceptron

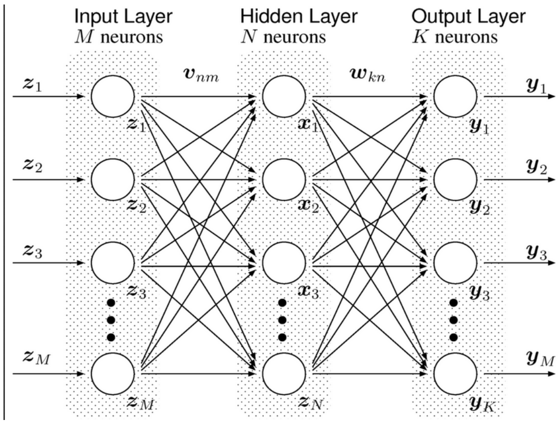


Fig 3.1

3.1 Simple neural network: In a simple artificial neural network, neurons act like switching units. Connections between neurons carry an activation signal of varying length. Typically, neurons are connected in layers where the signal reaches from the input layer to the output layer. The weights of the connections between the neurons are updated by backpropagation which is a technique by which the error or loss function is minimised by changing the magnitude of the interconnections between the neurons.

Each layer of the neural network has an activation function which converts neuron's weighted input into its output activation.

Each training instance in this case was a patch of size k*k*k where the value of k was repeatedly set to different values like 3, 5, 7 etc. Later it was found that 3*3*3 patch size performs the best as it captures only the most crucial local features near the midpoint of the cubical patch.

The patch could be rasterized to a linear array of size k*k*k. This array can be understood to be a set of features for a sample instance.

We design our artificial neural network in a way that based on these k*k*k features which represent the values of voxels in a cubical path, the model predicts the value of the middle voxel of this patch.

The following architecture was adopted for this ANN

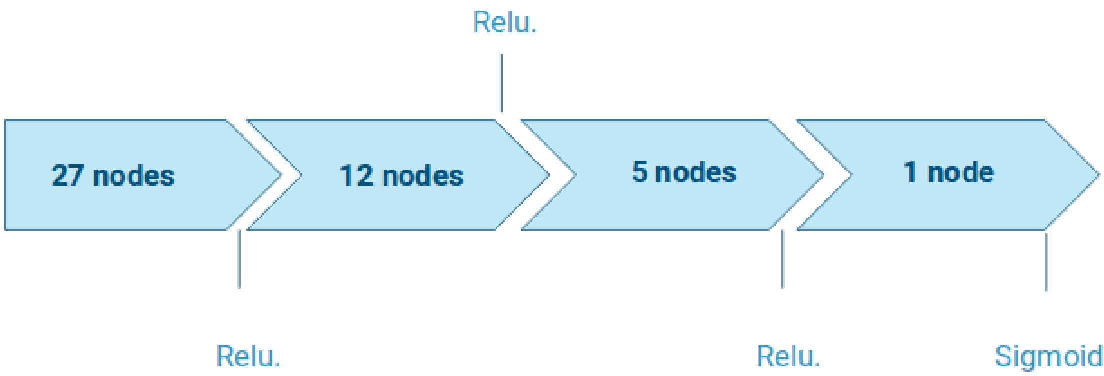


Fig 3.2

Since there are a whole lot of black patches, it is expected that the skewness in the data might bias the output of our learner. Hence, stratification of the data points is essential in such circumstances. Stratification involves maintaining the same kind of distribution of classes in the training samples as it is in the actual data.

The model was trained using 2,00,000 stratified patches from each 3D image.

Convolutional Neural Networks: Convolutional neural networks are similar to and were initially inspired from artificial neural networks. Although, the main difference lies in the fact that each layer of CNN has volumes of neurons with their weights expressed in the form of kernel matrices. Convnet make the assumption that the input is in the form of 2D or 3D images.

Convolutional neural networks are adept at classifying and regressing images.
The different variations in the architecture were tried. The number of layers were varied from 3, 4, 6, 8 to 11. The kernel size was initially kept at 3*3*3 and then increased to 5*5*5.

****

5-fold cross validation was performed on the data. This was done for both intra-specie as well as for inter-specie.
Following are the results which one obtains by feeding the images to a trained neural network.

****

Taiwan fly 1

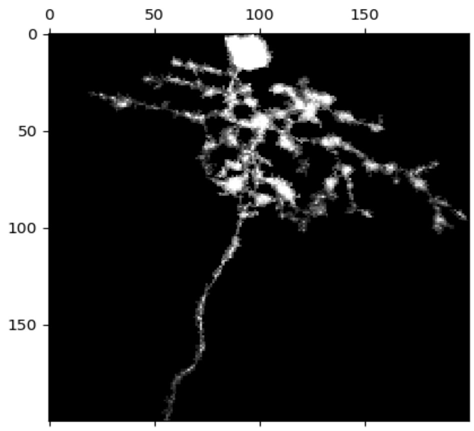


Fig 1.1

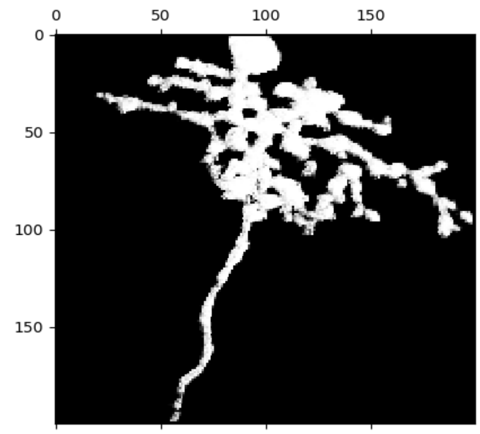


Fig 1.2

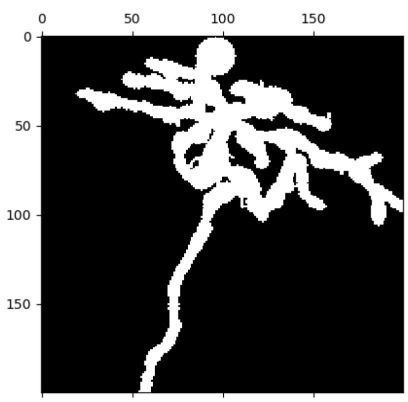


Fig 1.3

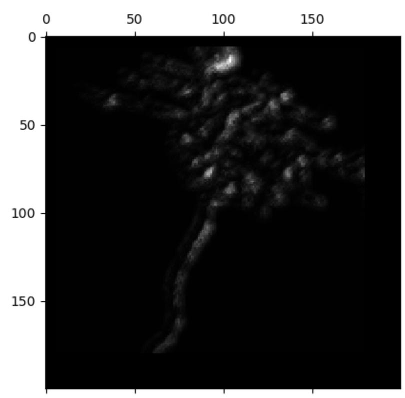


Fig 1.4

Taiwan fly 2

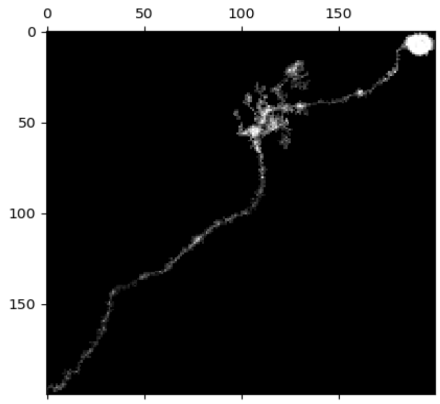


Fig 2.1

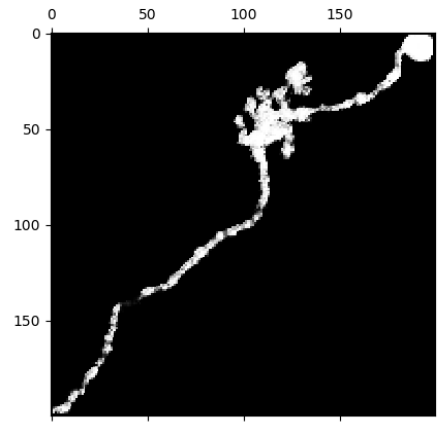


Fig 2.2

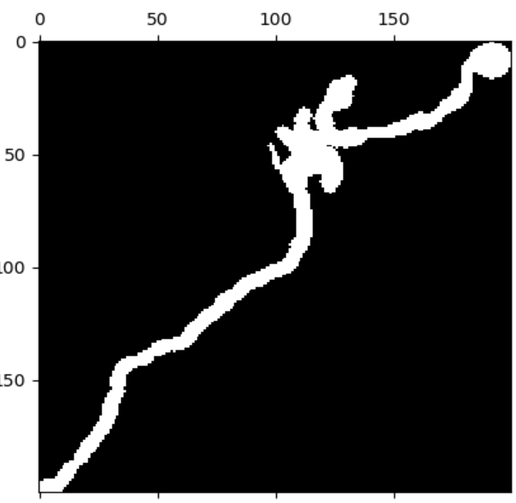


Fig 2.3

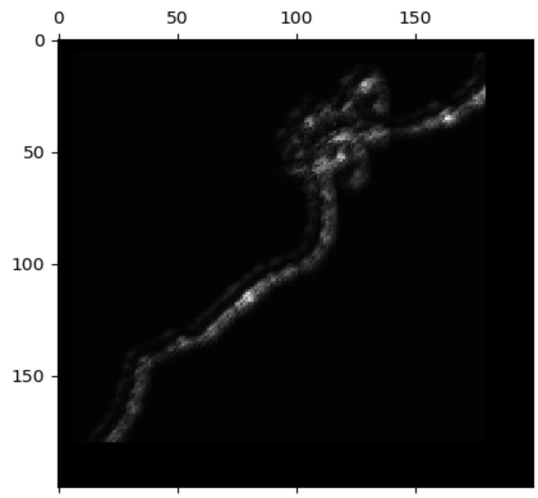


Fig 2.4

Taiwan fly 3

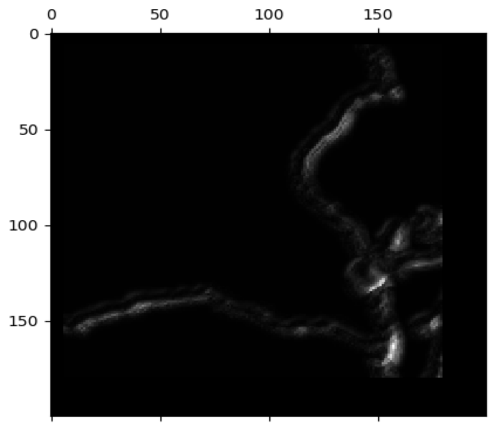


Fig 3.1

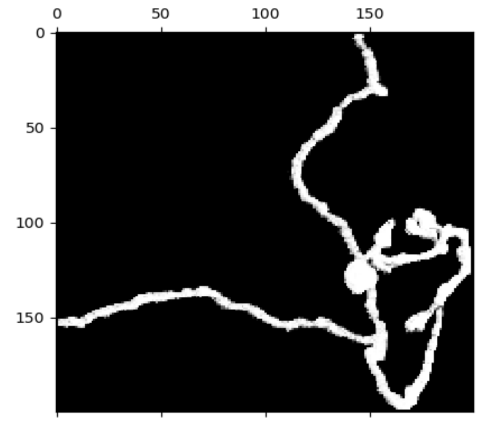


Fig 3.2

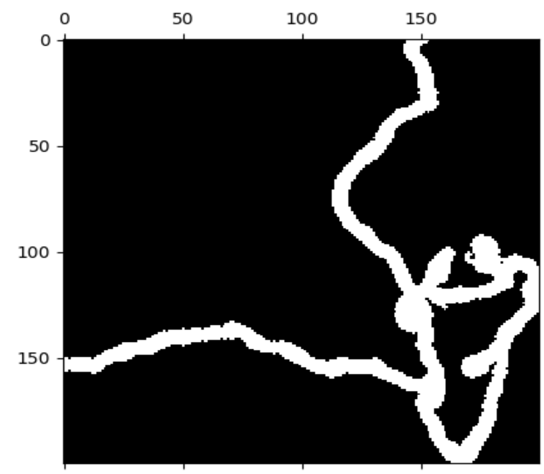


Fig 3.3

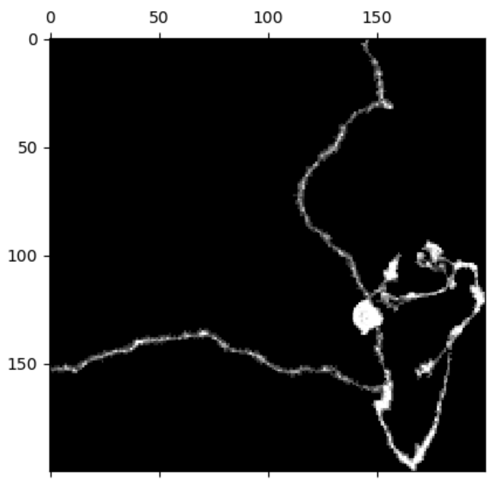


Fig 3.4

Janelia fly 7

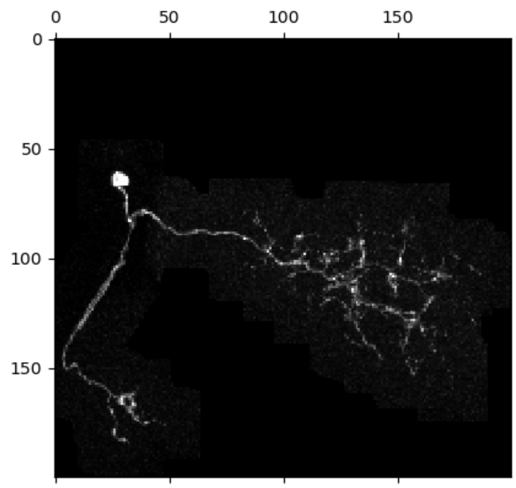


Fig 4.1
Original image

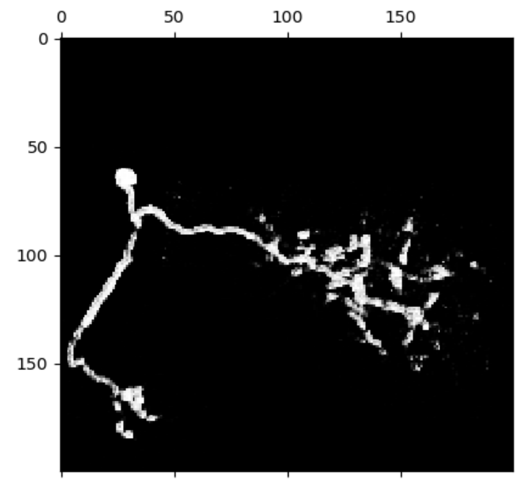


Fig 4.2
MLP reconstruction

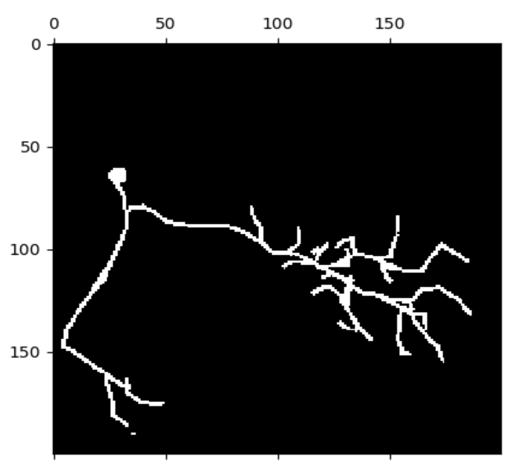


Fig 4.3
Ground truth

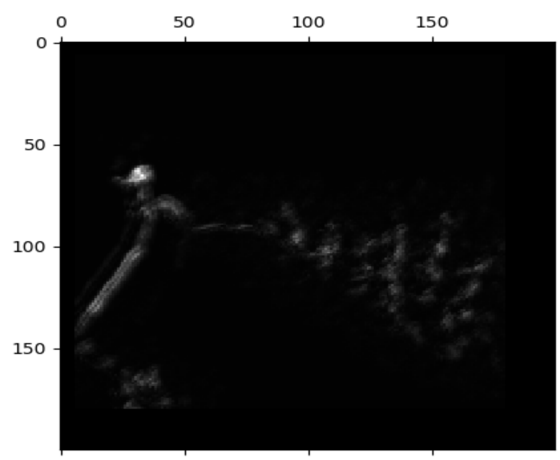


Fig 4.4
CNN reconstruction

We can clearly see here that the output of the ANN still has many discontinuities and breaks. This is handled by some post-processing measures which will be explained later in the report.

Chapter 4 : Quantitative Analysis of the results obtained

3.1 Fruit fly 7

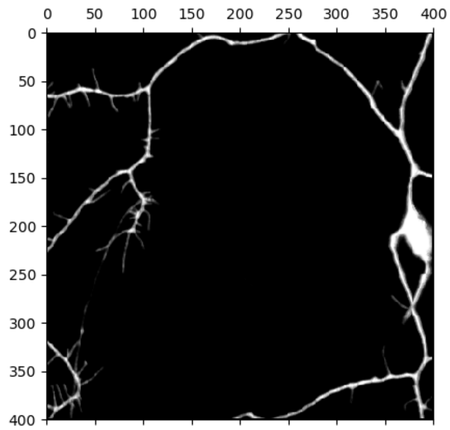


Fig 5.2
MLP Reconstruction

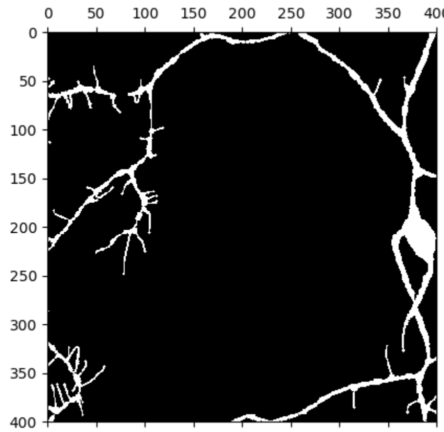


Fig 5.3
Ground truth

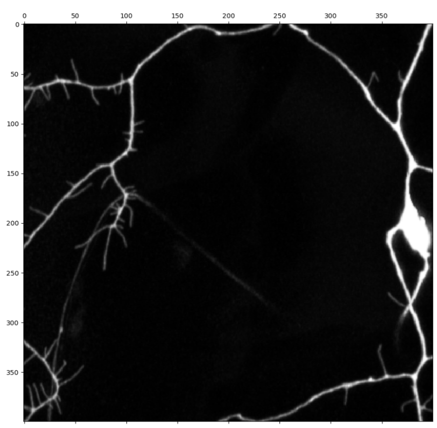


Fig 5.1
Original image

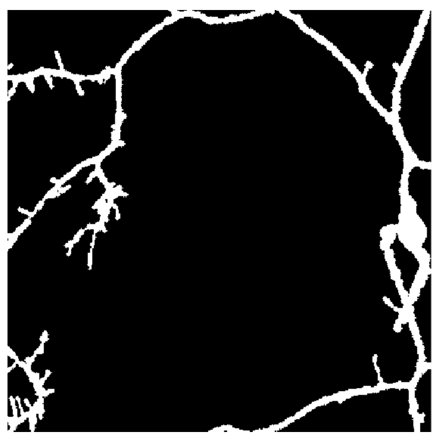


Fig 5.4
All path pruning

We can observe that the MLP reconstruction successfully prunes the strands which are there in the original microscopic image but are not there in the ground truth. In certain areas of doubt, where there are a lot of branches coming out from the main branch of the neuron, the MLP reconstruction retains those small outgrowths as we can see in the lower-left region of the reconstructed image. Another thing to notice is that the reconstruction given by the MLP is a little thinner than the ground truth and the one given by the all path pruning is a little thicker.

3.2 Statistics

Since our data has a lot of black patches, we don't want our output to be biased in favour of specificity over sensitivity. So we randomly choose 10,000 black and white patches from the image and analyse the accuracy.

Fig 21.1

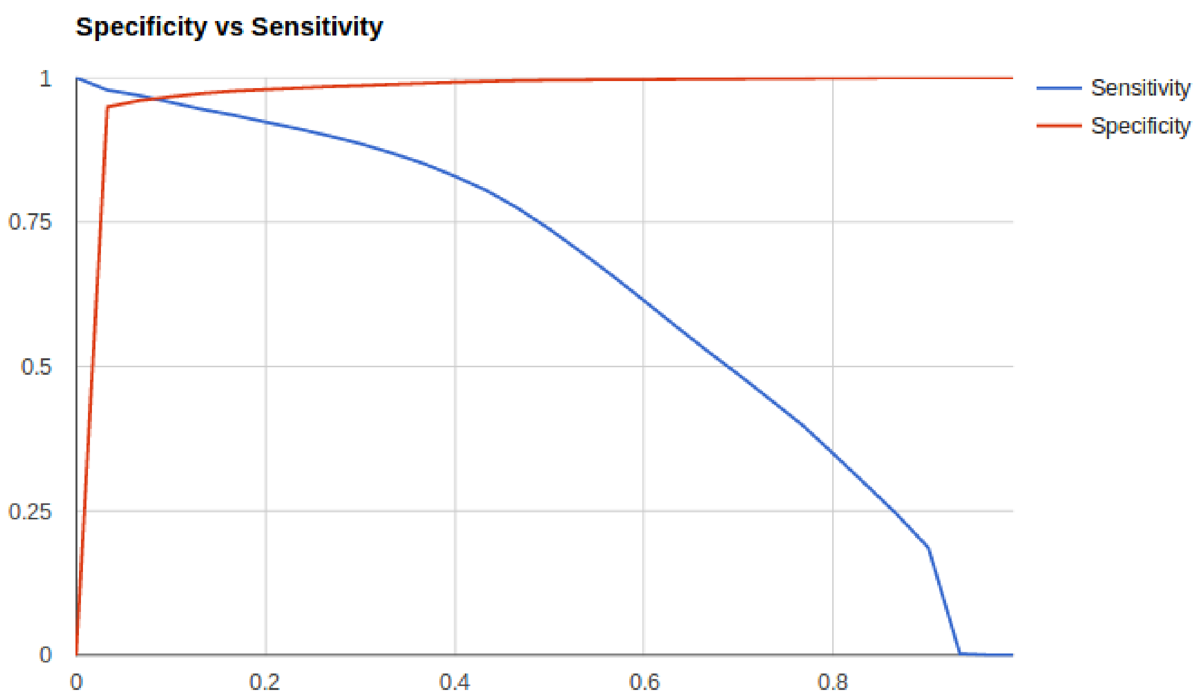


Fig 21.2 MLP

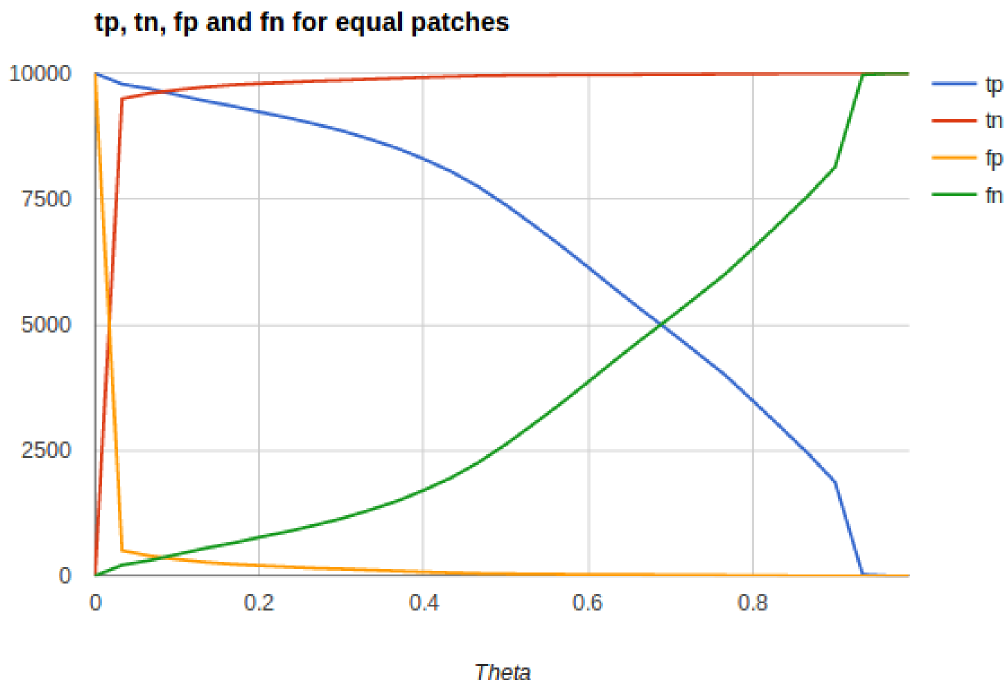

tp, tn, fp and fn for equal patches
10000
7500
5000
2500
0
0
0.2
0.4
0.6
0.8
Theta
tp
tn
fp
fn

Fig 21.3

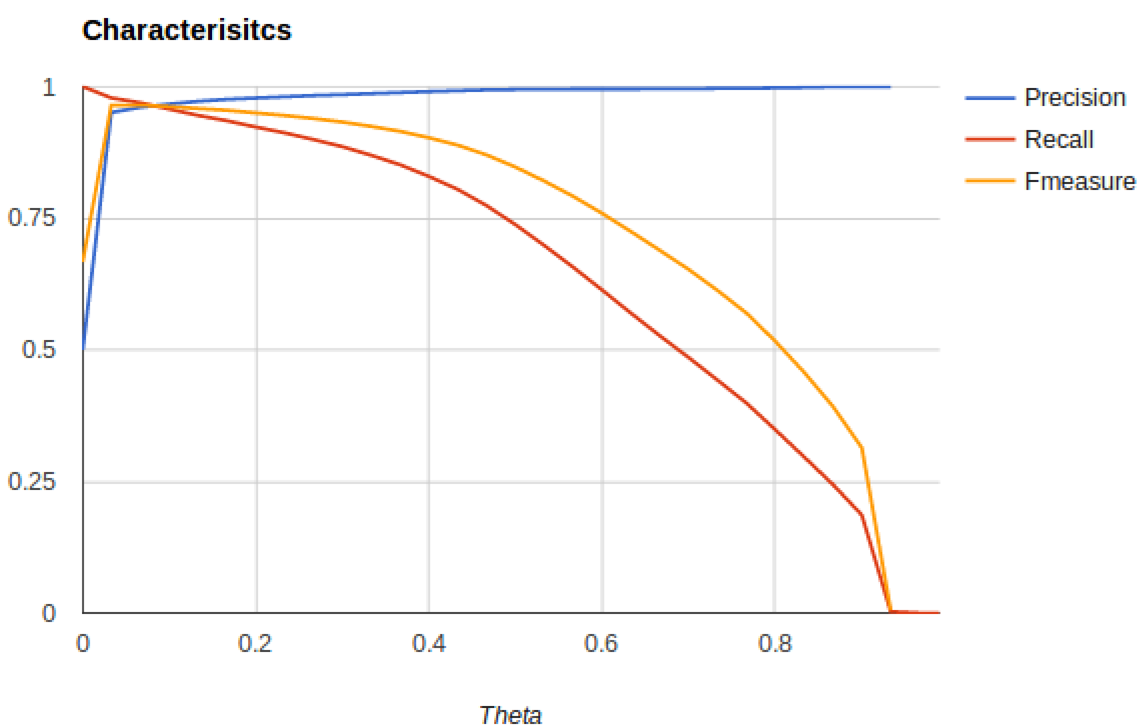

Characterisitcs
1
0.75
0.5
0.25
0
0
0.2
0.4
0.6
0.8
Theta
Precision
Recall
Fmeasure

Fig 21.4

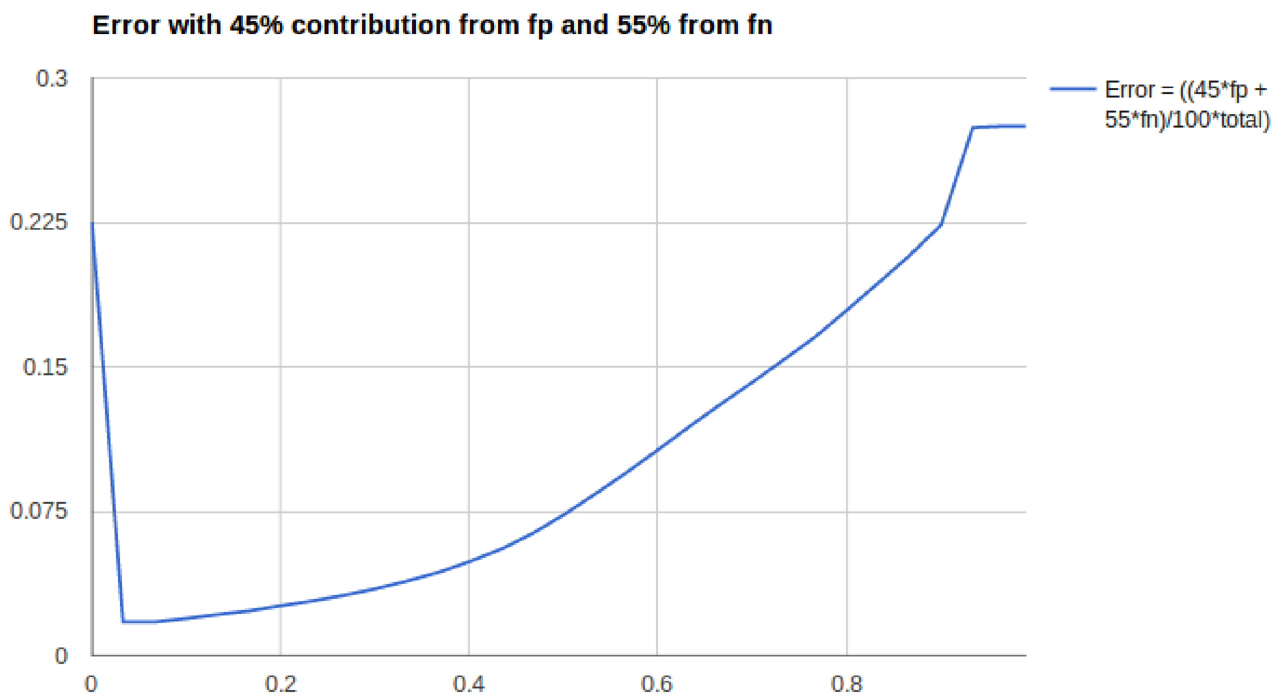
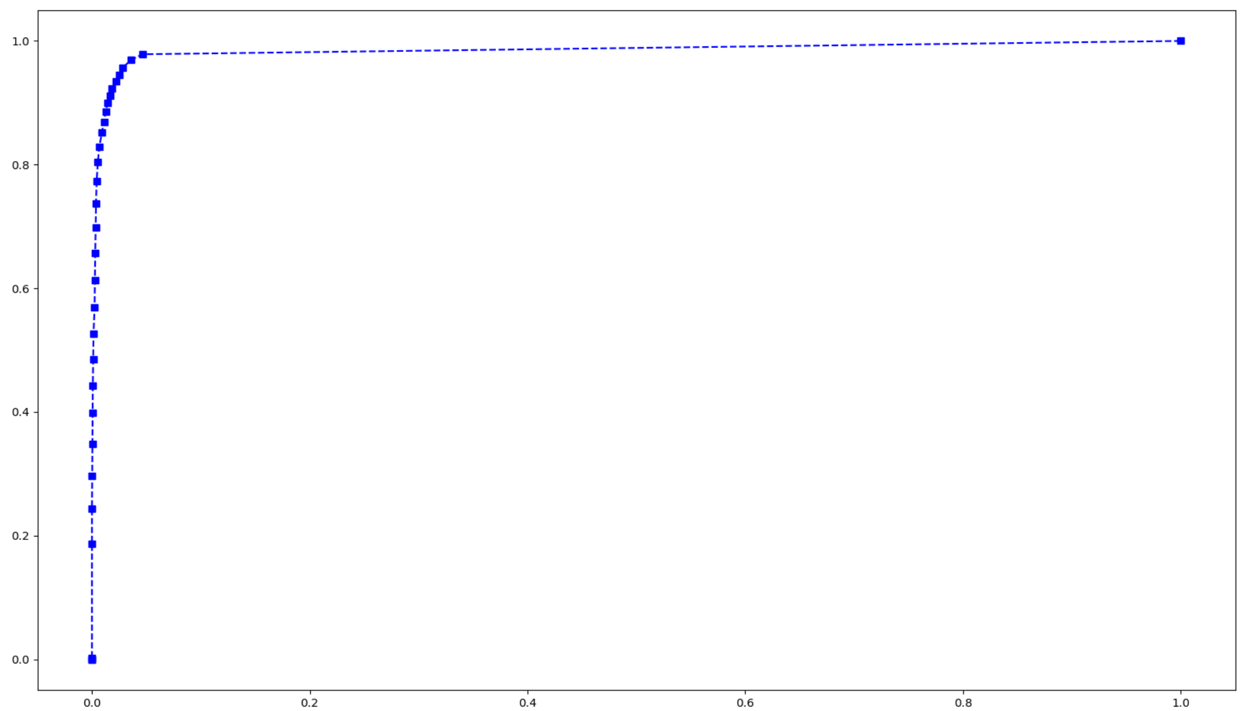


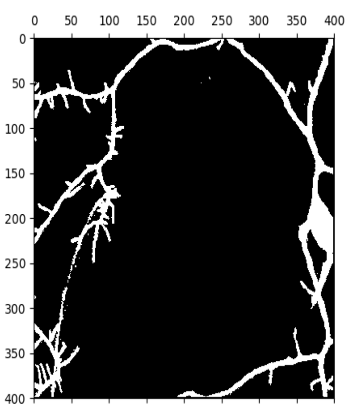
Fig 21.5



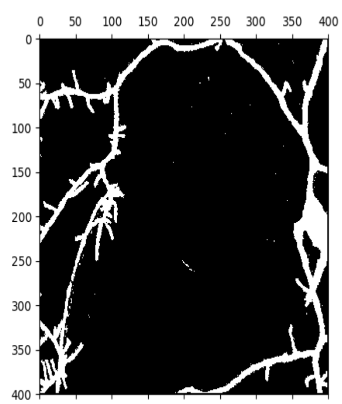
ROC curve fruit fly

Fig 21.6

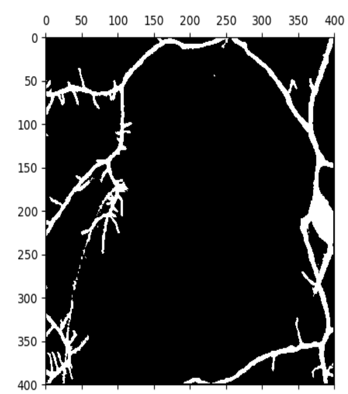
3.3 Effect of thresholding on the neuron structure



0.1



0.2



0.3

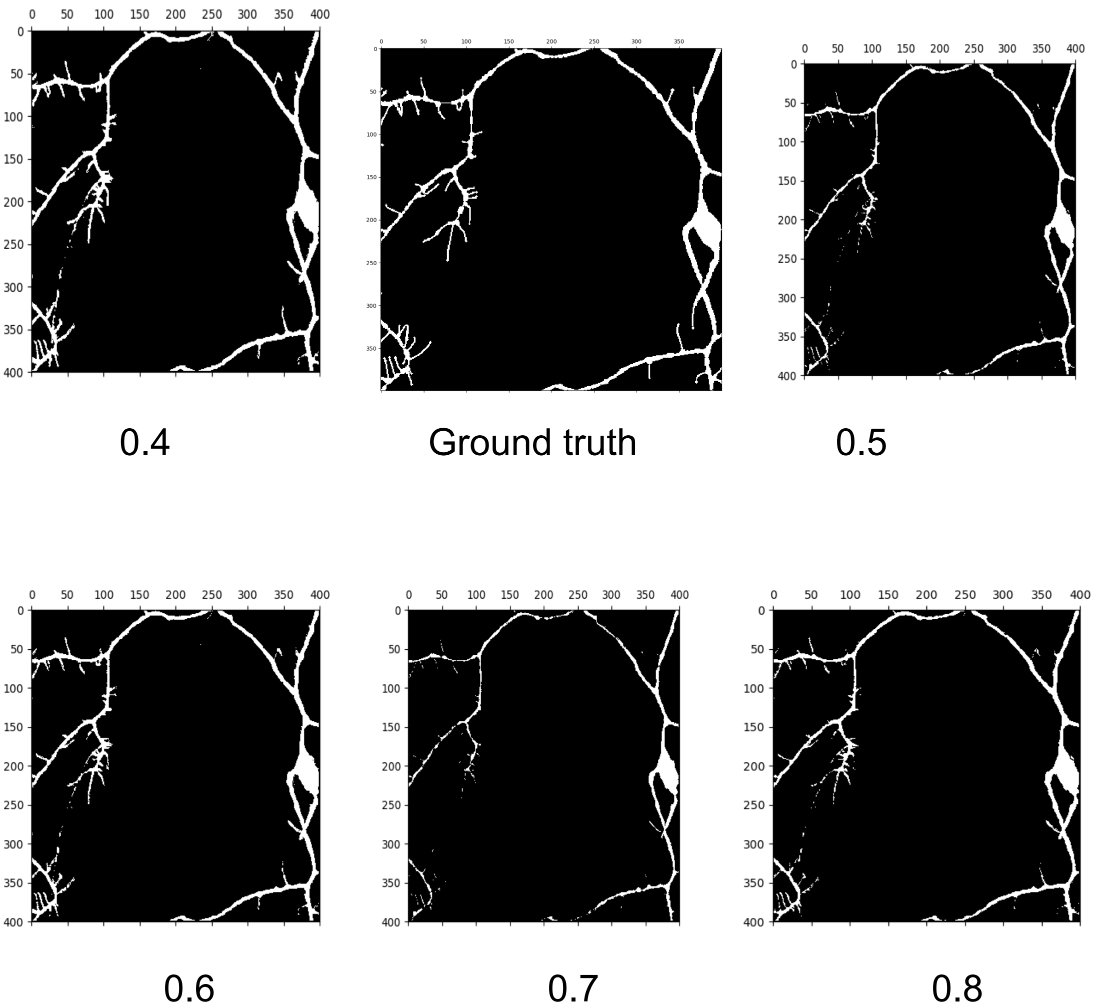


Fig 22.1

One can observe that on increasing the threshold, the reconstruction become more pruned, noise gets reduced. But since false positives are to be penalised less than false negatives, we will choose a smaller threshold.

Inter specie vs Intra specie 5 cross validation

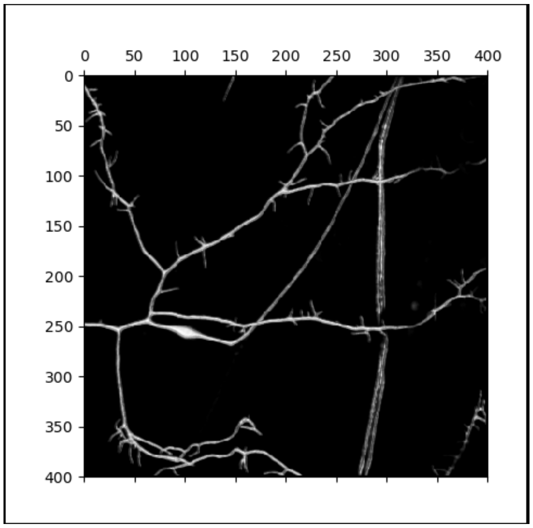


Fig 23.1

Intra specie reconstruction

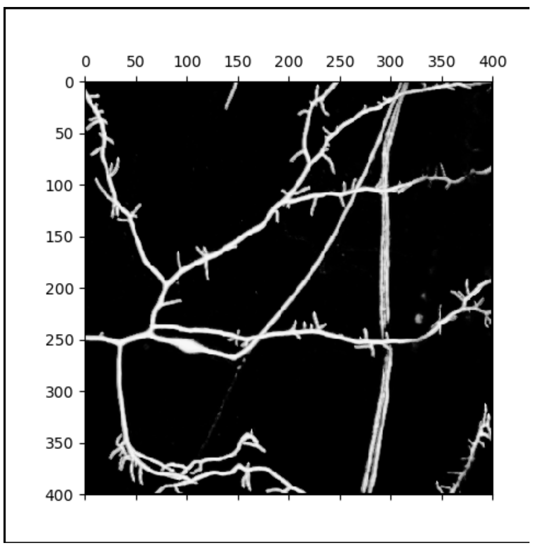


Fig 23.2

Inter-specie reconstruction

One can observe that the inter specie reconstruction appears more thick and over-constructed than the intra specie neuron reconstruction. This might be because of the fact that most of the other species are thicker in their anatomy than that of fruitfly and the learner after seeing their patches might have got biased.

Intra specie cross validation produces a reconstruction which is more pruned and has less noise and other unnecessary artefacts which are not there in the ground truth.

Chapter 5 Overcoming the limitations

4.1 Entangled dendrites

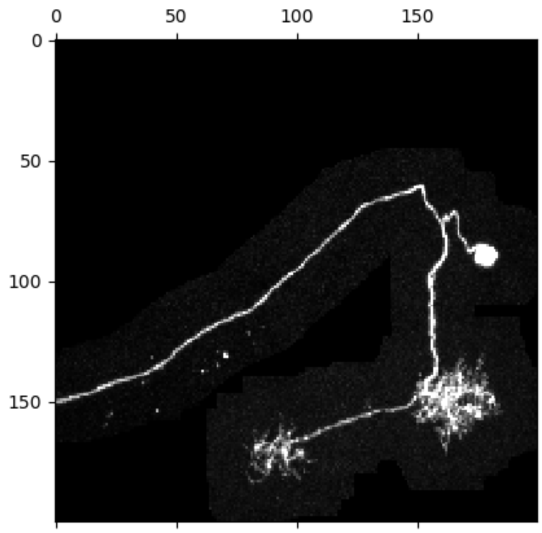


Fig 24.1

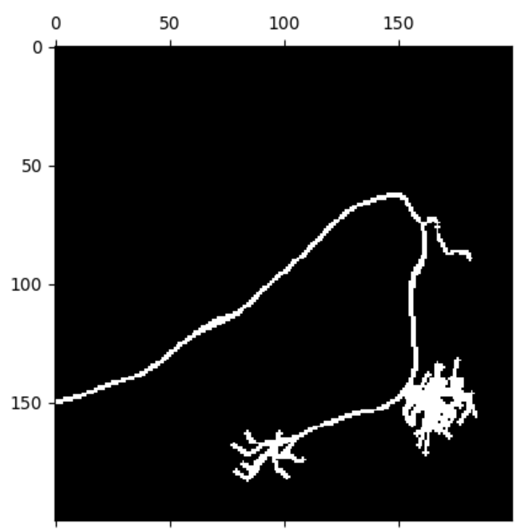


Fig 24.2

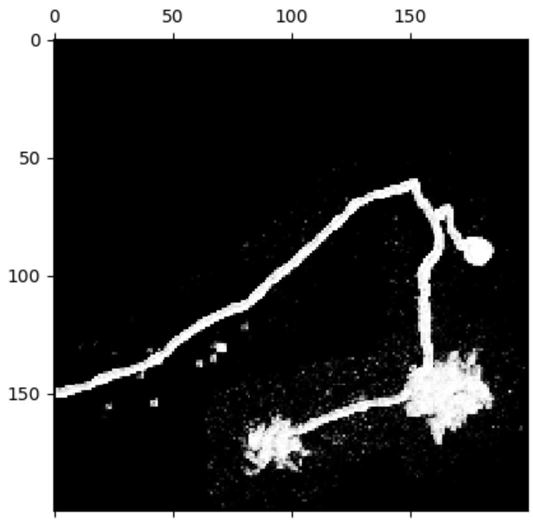


Fig 24.3

In certain images, where there are portions within an image with huge amount of entanglement, the neural network model produces a homogenous blob like structure which is visible in the above Fig 24.3

4.2 Discontinuities

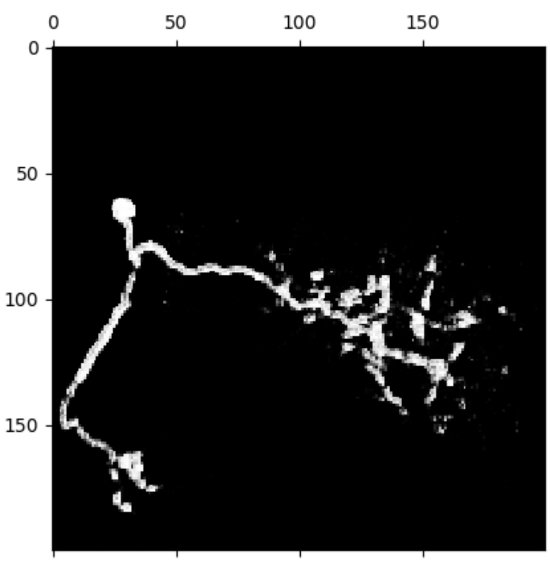


Fig 25.1

One very clearly visible problem as we see here is that of discontinuities in the neuron reconstruction. To overcome this problem, certain post processing techniques have been proposed. Morphological operations on images have been used as a measure to reduce the discontinuities.

Opening: One particular measure which has been used in this project is that of opening. Opening stands for two operations in succession erosion and dilation. As seen in this particular case, opening first removes small speckles of noise which are there in the image while preserving the big blobs and then it dilates them to bridge the gaps in the neuron structure. A cubical structuring element was used with edge 3 pixels long.

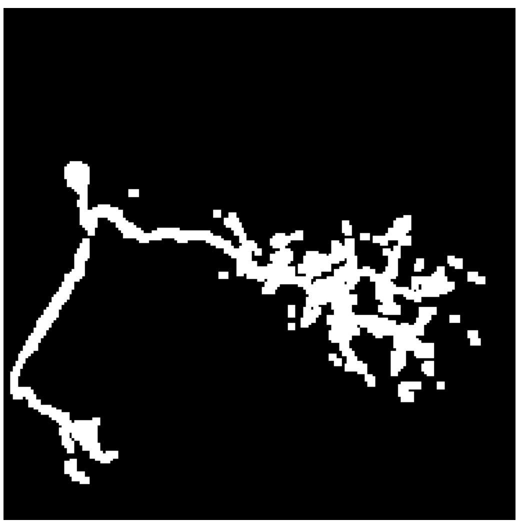


Fig 25.2

Chapter 5
Conclusions:

From the results obtained over the course of this project, we can conclude that artificial neural network based learner performs comparable or slightly better than the best neuron tracing algorithms used currently except in certain species like chicken and humans where the signal to noise ratio is extremely low. But apart from that the accuracy obtained on an average by choosing a global threshold for all the images comes out to be 98% for all path pruning algorithm it comes out to be 96.5%. The accuracy computation penalises false negatives a little more than false positives (55:45).

The challenge next would be to identify as to what can be done so that the model can work equally well in reconstructing human and chicken neurons.

One reason for this can be attributed to less number of image samples available for training the patches of these neurons species, hardly two for each. Furthermore, each image sample has a considerably high noise content.

The outputs of the CNN also appear to be not encouraging despite that Convnets are more suited for classifying images. Again, this can be attributed to less amount to training data. Since CNN is a deep network, it requires a lot more training samples for learning the features of images. More work on using CNNs to reconstruct neurons can be done once the bottleneck of data cruch is resolved.

Post processing techniques like opening help bridge the gaps in the neuron structure well. The converse of this which is called closing or dilation followed by erosion was experimented with. This led to over-emphasizing of the noise and led to highly blog like structure. However, morphological operations are open to customization in terms of shape of structuring element, size in terms of voxels of the structuring element.

Appendix 1

References

1. Automatic 3D neuron tracing using all-path pruning
   https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3117353/
   Published online 14th June 2011
   Author : Hanchuan Peng, Fuhui Long, Gene Myers

2. A Broadly Applicable 3-D Neuron Tracing Method Based on Open-Curve Snake
   https://www.researchgate.net/publication/50377555_A_Broadly_Applicabl

e_3-D_Neuron_Tracing_Method_Based_on_Open-Curve_Snake
Published online: March 2011

Authors: Yu Wang, Arunachalam Narayanaswamy, Chia-Ling Tsai, Badrinath Roysam

3. TReeMAP: Automatic 3D Neuron Reconstruction Based on Tracing, Reverse Mapping and Assembling of 2D Projections
http://home.penglab.com/papersall/docpdf/2015_Neuroinfo_TReMAP.pdf
Authors: Zhi Zhou, Xiaoxiao Liu, Brian Long and Hanchuan Peng
Published online: 2015
4. Digital Image Processing by Rafael C. Gonzalez and Richard E. Woods. Third Edition

5. Introduction to Machine Learning Ethem Alpaydin Second Edition

6. https://www.nitrc.org/forum/
This is a forum to discuss any issue about the the neuron data, Vaa3D the neuron processing application and related things.

7. Keras: Deep Learning library for Theano and TensorFlow
https://keras.io/

8. Stack overflow
http://stackoverflow.com/